Summer 8-2020

A Novel Correction for the Adjusted Box-Pierce Test — New Risk Factors for Emergency Department Return Visits within 72 hours for Children with Respiratory Conditions — General Pediatric Model for Understanding and Predicting Prolonged Length of Stay

Sidy Danioko
Chapman University, danio100@mail.chapman.edu

A Novel Correction for the Adjusted Box-Pierce Test — New Risk Factors for Emergency Department Return Visits within 72 hours for Children with Respiratory Conditions — General Pediatric Model for Understanding and Predicting Prolonged Length of Stay

A Dissertation by

Sidy Danioko

Chapman University

Orange, CA

Schmid College of Science and Technology

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Computational and Data Sciences

August 2020

Committee in charge:

Cyril Rakovski, Ph.D., Chair

Daniel Alpay, Ph.D.

Louis Ehwerhemuepha, Ph.D.

Gary Doran, Ph.D.

Edward McFowland III, Ph.D.

The dissertation of Sidy Danioko is approved.

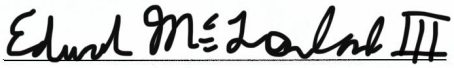
Cyrl Rakovski

Cyril Rakovski, Ph.D., Chair

Daniel Alpay, Ph.D.

Louis Ehwerhemuepha, Ph.D.

Gary Doran, Ph.D.



Edward McFowland III, Ph.D.

August 2020

A Novel Correction for the Adjusted Box-Pierce Test — New Risk Factors for Emergency Department Return Visits within 72 hours for Children with Respiratory Conditions — General Pediatric Model for Understanding and Predicting Prolonged Length of Stay

by Sidy Danioko

ACKNOWLEDGMENTS

This thesis work has taken me immeasurable time and energy than I had ever thought, and during this adventure it was frequent for me to loose hope and confidence. Without the supports from advisors, professors, and friends, this research project would not have been a tangible thing. I wish to express my sincere and humble gratitude to Dr.Rakovski who walked me into this gigantic adventure, by constantly giving me abundant help, offering gigantic assistance, unweaving soft and strict guidance. He was reviewing the writing, as when I advanced in my work, in an atmosphere of friendship, conversation and categorical rejection of easy and wrong analysis, mostly when there is a more rigorous and logical way to explain complex mathematical, computational, statistical concepts and data related issues. My respectful and unique thanks to Dr. El-Askary, current Director of the Computational and Data Science(CADS) Department for providing me with the financial and technological means, which produce substantial work. In some sense, I can say that Dr. El-Askary is symptomatic of this big achievement for not only believing in me since day one, but also for being on my side during the most difficult circumstances. Special thanks to CADS's faculty members for making me ready for current and future challenges through their killer homework and exams. I thank Dr. Daniel Alpay(Chapman University), Dr. Louis Ehwerhemuepha (CHOC Children's Health System), Dr. Gary Doran(Jet Propulsion Laboratory), and Dr. Edward McFowland III(University of Minnessota) for accepting to be part of my doctoral committee. Parallely, I would like to thank all my groupmates Eshan Yaghmaei, Alex Barrett, Jianwei Zheng(Arnold), and Chris Watkins who were endlessly answering to my calls during the most stressing periods of this thesis work. I would like to also thank my friend Ismael De Paiva for his considerable contribution to this achievement. Finally, I highly appreciate the acknowledgment and tribute paid to my best friend Kyle Anderson for his unyielding moral and tangible supports to me when the road to this achievement was hopeless.

LIST OF PUBLICATIONS

During my sejourn at chapman university, I have worked on a battery of projects lying at the intersection of different research areas and in connivance with my education background and future ambitions. To the date of the submission of my doctoral thesis, the projects that are either submitted to top tier journals or under preparation for submission are:

- A Novel Correction for the Adjusted Box-Pierce Test.
- Risk Factors for Emergency Department Return Visits within 72 Hours for Children with Respiratory Conditions.
- General Pediatric Models for Understanding and Predicting Prolonged Hospital Length of Stay.
- Computational investments Strategies via the Use of Multivariate Time Series, Machine Learning and Deep Learning Models and Stock Selection.
- Effects of Imputation Methods on the Accuracy of Predictive Models.
- Studies of New Kinds of Stochastic Processes.

ABSTRACT

A Novel Correction for the Adjusted Box-Pierce Test — New Risk Factors for Emergency Department Return Visits within 72 hours for Children with Respiratory Conditions — General Pediatric Model for Understanding and Predicting Prolonged Length of Stay

by Sidy Danioko

This thesis represents the results of three research projects that underline the breadth and depth of my interests.

Firstly, I devoted some efforts to the well-known Box-Pierce goodness-of-fit tests for time series models which has been an important research topic over the last few decades. All previously proposed tests are focused on changes of the test statistics. Instead, I adopted a different approach that takes the best performing test and modifying the rejection region. Thus, I developed a semiparametric correction of the Adjusted Box-Pierce test that attains the best I error rates for all sample sizes and lags and outperforms all previous global time series goodness-of-fit approaches.

Secondly, I aimed to study and identify novel risk factors significantly associated with 72-hour return visits to emergency departments. We queried data consisting of 185,000 ED visits of patients less than 18 years in the United States using the Cerner® Health Facts Database. A nested mixed-effects logistic regression model to provide statistical inference on associated risk factors was built, and a representative set of machine learning algorithms for our predictive modeling task was selected. New respiratory conditions including acute bronchiolitis, pneumonia, and asthma were identified as risk factors for return visits to ED.

Thirdly, I ambitioned to design and implement a comprehensive study to identify new clinical and demographic factors associated with prolonged length of stay (> two weeks) among pediatric patients (aged 18 years and under) in a number of free-standing pediatric and mixed

medical facilities. We implemented a mixed effect model to assess the statistical significance and effect sizes of age, race/ethnicity, number of medications, medical family history, presence of infection agents (fungi, bacteria, virus), cancer diagnoses, and other conditions as well as some clinical variables. A stochastic gradient model was also implemented for prediction. From the mixed-effects model, 11 main effect predictors were found to be significantly and statistically associated with an increase in the odds of prolonged length of stay. The area under the operator characteristic curve (AUROC) for the mixed-effects model was 0.887 (0.885, 0.889) and the extreme gradient boosting model attained an AUROC of 0.931 (0.930, 0.933).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CHOC	Children Hospital of Orange County
HV	Human Immunodeficiency Virus
HIPAA	Health Insurance Portability and Accountability Act
ICD	International Classification of Disease
WHO	World Health Organization

LIST OF SYMBOLS

Q^a_{BP}	Adjusted Box-Pierce
B^k	Backschift operator
Q_{BP}	Box-Pierce
χ^2	Chi Square
$N_n(0, I)$	Error(white noise) distribution
$\hat{r}_k$	Estimated autocorrelation coefficient at k
Q_{LB}	Ljung-Box
Q_M	Portmanteau Test
$\tilde{R}_m$	Peña and Rodríguez
$\tilde{Q}_{WL}$	Weighted McLeod and Li's Test
$\tilde{Q}_{WM}$	Weighted McLeod and Li's Test

1 Introduction

Currently, data accumulation is accelerating and touching every domain of life. For example, in physics time series arises quite often when studying very dynamic complex systems. In engineering, electricians are always engaged in better understanding the time dependent aspects of power flow over a fixed interval of time. In medicine, doctors daily or weekly conduct lab tests or other screening techniques on patients. In social sciences, the population growth rates are regularly measured in the hope of prescribing trends and design recommendations for the future. In finance and economics, the daily, weekly, and monthly prices of stocks are constantly collected for studies leading to better investment plans. In the industry world, some scientists observe the time evolution of the densities of plasma.

Data come in many forms including structured, unstructured, semi-structured, discrete, continuous, high dimensional, to just name a few. Due to the existence of various types of data and the multitude of research questions that can be posed, understanding and modeling have been attracting various communities of researchers and practitioners. Particularly important has been the presence of the notion of data and the science behind data (data science) at the intersection of all the above-cited data related activities. Arguably, data science has been viewed as a study domain that lies at the intersection of math and statistical knowledge, hacking skills, and very broad and deep expertise. From all the difficulties of generating a considerable amount of data to a world submerged by data, the field of data science has evolved in several directions and matured in numerous aspects. With time and the abundance of data (structured and unstructured), data science has gradually become one of the intellectually stimulating research areas within companies, universities, and governments. In fact, finding sound mathematical theory and computational algorithms have produced unprecedented success stories. Nevertheless, uncountable practical applications and active research topics continue to emerge. Thus, it is still acceptable to say that data science has

not yet reached a completely mature stage.

I have devoted parts of my doctoral work to both extending theoretical results and investigating medical data in the hope of gaining understanding, insight, and knowledge with respect to situations that are inflecting serious psychological and financial burdens on patients and hospitals as well. This dissertation consists of four parts. Part I is devoted to the introduction. Part II is focused to creating a novel correction to the adjusted version of the Ljung-Box statistic, one of the most popular time series goodness-of-fit diagnostic test statistics. Part III addresses the risk factors of 72-hour return visits attributable to the most common respiratory conditions and the contribution of non-respiratory comorbid conditions/diseases. Part IV is concerned with designing and implementing a comprehensive study to identify new clinical and demographic factors associated with prolonged length of stay ($>$ two weeks) among pediatric patients (aged 18 years and under) in a number of free-standing pediatric and mixed medical facilities.

2 A Novel Correction for the Adjusted Box-Pierce Test

The Box-Jenkins algorithm is a general systematic approach for model checking of a time series model. Examples of the approach can be found in [1], [2] and , [3]. A well-fitting model produces residuals that are free of correlation. Thus standard goodness-of-fit approaches are in essence global tests for absence of correlation among estimated residuals. Accordingly, many statistical techniques have been designed to assess the absence of correlation among the time series model residuals.

Following classical notation, let $\{X_t\}$ be an observed time series generated by a stationary and invertible ARMA(p,q) process $\phi(B)X_t = \theta(B)\epsilon_t$, where $\phi(B)$ and $\theta(B)$ are the autoregressive and moving average characteristic polynomial and $B^k X_t = X_{t-k}$ is the backshift operator. The desired parameters, ϕ_i and θ_i are estimated using maximum likelihood or least squares methods to obtain $\hat{\phi}_i$ and $\hat{\theta}_i$, the residuals are calculated via $\hat{\epsilon}_t = \hat{\theta}^{-1}(B)\hat{\phi}(B)X_t$ and the sample auto-correlation coefficients are in turn obtained from $\hat{r}_k = \sum_{t=k+1}^{n} \hat{\epsilon}_t\hat{\epsilon}_{t-k}/\sum_{t=1}^{n} \hat{\epsilon}_t^2$.

In recent years, many techniques have been employed to test the global hypothesis of all autocorrelations up to a certain lag, $\mathbf{H}_0 : r_1 = r_2 = ... = r_m = 0$. In general, these techniques are designed as weighted sums of squares of the estimated autocorrelations and they can produce misleading conclusions due to deviations from the asymptotic limiting distribution in moderate size samples [4] , [5], [6]. Thus, a new and more robust test is proposed in this research that attains precise type I error rates for all sample sizes.

The history of portmanteau tests traces its roots back to the Box-Pierce diagnostic test defined as [6], [7] :

$$Q_{BP} = n \sum_{k=1}^{m} \hat{r}_k^2, \tag{2.1}$$

where n, m, and $\hat{r}_k$ represent the sample size, number of lags being tested and the sample auto-correlation of order k of the residuals respectively. The authors showed that the asymptotic distribution of Q_{BP} is approximately $\chi^2(m\text{-}p\text{-}q)$ but considerable deviations for moderate sample sizes have been observed [7], [8], [9]. That deficiency entails imperfections of type I error rates and prompted the design of a weighted and improved versions of the test. In their stimulation studies, Ray and Xiaolou focused on investigating the type I errors in the χ^2_m setting [4]. They remarked that Box-Pierce rejects too often because of the fact that the test statistic is too small.

Ljung and Box were the first ones to propose a design that assigns larger weights to residuals estimated with more data [7]:

$$Q_{LB} = n(n+2)\sum_{k=1}^{m}\frac{\hat{r}_k^2}{n-k} = n\sum_{k=1}^{m}\frac{n+2}{n-k}\hat{r}_k^2. \tag{2.2}$$

The Box-Pierce and Ljung-Box tests are asymptotically equivalent. The Ljung-Box test has been shown to overcorrect in moderate samples [4]. They also realized that Ljung-Box inflates the test statistic using a variance estimate of the residuals. They further showed that on moderate sized data, Q_{LB} rejects too often because the test statistic is too small.

Li and McLeod refined the Q_{BP} test [9] by proposing the following statistic,

$$Q_{LB} = Q_{BP} + \frac{m(m+1)}{2n} = \frac{m(m+1)}{2n} + n\sum_{k=1}^{m}\hat{r}_k, \tag{2.3}$$

This approach only corrects the mean of the Box-Pierce statistic and consequently fails to properly adjust the type I error rates.

Monti proposed a portmanteau test based on the residual partial autocorrelations [10]. The

test is defined as,

$$Q_M = n(n+2)\sum_{k=1}^{m}\frac{\hat{\pi}_k^2}{n-k}, \tag{2.4}$$

Monti showed via simulations that the performance of Q_M is comparable to that Q_{LB} [10]. In addition, he concluded that in certain scenarios, Q_{LB} outperfroms Q_M.

Peña and Rodríguez proposed a test based on a different measure of dependence of the residual autocorrelations, [11],

$$D = n(1 - |\hat{R}_m|^{1/m}), \tag{2.5}$$

where

$$\tilde{R}_m = \begin{pmatrix} 1 & \hat{r}_1 & \dots & \hat{r}_m \\ \hat{r}_1 & 1 & \dots & \hat{r}_{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{r}_m & \hat{r}_{m-1} & \dots & 1 \end{pmatrix} \tag{2.6}$$

In their work, the authors showed that under particular conditions, their test greatly outperformed Q_{LB} test. Furthermore, they demonstrated that the test had an advantage over the McLeod and Li's test regardless of sample size. However, the convergence of the asymptotic distribution of the test developed by Peña and Rodríguez is very slow [12].

Fisher proposed new weighted versions of the Box-Pierce and Monti's tests, the Q statistic: [5],

$$\tilde{Q}_{WL} = n(n+2)\sum_{k=1}^{m}\frac{m-k+1}{m(n-k)}\hat{r}_k^2, \tag{2.7}$$

and

$$\tilde{Q}_{WM} = n(n+2)\sum_{k=1}^{m}\frac{m-k+1}{m(n-k)}\hat{\pi}_k^2, \tag{2.8}$$

A comparison simulation study by [13] showed that for small sample size and m values Q_{WL} performs better than Q_{LB}. For moderate sample sized data, they also found that Q_{WL} does better than Q_{LB} and Q_{WM} outperforms Q_M.

To remedy some of the shortcomings of all previously existing tests, Kan and Wang proposed a new modification of the portmanteau test, widely called the adjusted Box-Pierce test [4]. They defined their statistic as,

$$Q_{BP}^{a} = m + \sqrt{\frac{2m}{Var[Q_{BP}]}}(Q_{BP} - E[Q_{BP}]), \tag{2.9}$$

The authors conducted an evaluation of various tests including Box-Pierce and Ljung-Box. The design of the adjusted Box-Pierce statistic (2.9) explicitly recenters and rescales Q_{BP} to attain the mean and variance of a $\chi^2(m)$ variable. The authors showed through simulations that the test possesses very good adherence to nominal type I error rates. In their comparison study, they found that both the distributions of Q_{BP} and Q_{LB} deviate from the expected variance of $\chi^2(m)$ distribution for small and moderate sample sizes and almost all choices for the value of m.

All of the above-mentioned tests exhibit deviations from the nominal type I error rates that compromise their performance. Thus, a new approach is proposed which aims at correcting the rejection region instead of redesigning the test statistic itself. This technique was introduced by Bernard in his effort to construct a more powerful alternative to Fisher's exact test [14] [15] and later by Boschloo [16]. The same idea of rejection region correction has been recently employed by Ehwerhemuepha et al to produce the best performing test for homogeneity for multinational distributions [17].

2.1 Methods

A model based correction of the rejection region of the adjusted Box-Pierce test was designed. A large scale simulation study was then conducted to not only estimate the correction, but to also assess the performance advantages (defined as adherence to the nominal type I error rates for all scenarios) of the proposed corrected method.

2.1.1 Simulation Study

For sample size values of $n = 40, 50, \ldots, 300$, we simulated 10^6 white noise samples, $s_{n1}, s_{n2}, \ldots, s_{n10^6} \sim N_n(0, I)$. These mimic the behavior of residuals of a well-fitting time series model (under the null). Next, the adjusted Box-Pierce test was applied to every sample and for all possible lags, m $(2 \leq m \leq n-1)$ and the corresponding p-values, $p_{nm1}, p_{nm2}, \ldots, p_{nm10^6}$ were obtained. For each pair (n, m), the estimated the type I error rate of the adjusted Box-Pierce test at alpha level of 0.05 was empirically estimated by $P^{n,m}_{\alpha=0.05} = \sum_{i=1}^{10^6} I\{p_{nmi} < 0.05\}/10^6$. Thus, for each sample size n, $n-2$ empirically estimated type I error rates yielding a dataset with three columns, n, m, and $P^{n,m}_{\alpha=0.05}$. Further, these datasets obtained from all individual sample sizes n were stacked to get an aggregated dataset with number of rows $\sum_{n=4}^{30} 10n(10n-2) = 934,920$.

2.1.2 Linear model

The primary idea of this study was to provide a model-based correction to the rejection region of the adjusted Box-Pierce test in order to attain improved type I error rates for all sample sizes and lags. We created six linear regression models trained on the simulated data described in the section above. These six models were trained on different subsets of the data split into sample size intervals, $(0, 50]$,$[51, 70]$, $[71, 90]$, $[91, 120]$, $[121, 200]$, and $[201, 300]$.

The difference in the type I error rate patterns for distinct sample seizes (shown in Figure 1) necessitated the use of separate models to achieve the desired level of fit. These linear models are complex as they encompass different powers of n, m and their 2-way interactions. The general formula adopted for the models was,

$$\begin{aligned} Y - 0.05 = \alpha_1 n^s + \alpha_2 m^p + \alpha_3 (n^s * m^p) + \alpha_4 (n^{2s} * m^{2p}) + \alpha_5 n^{2s} \\ + \alpha_6 (n^{3s} * m^{2p}) + \alpha_7 (n^{3s} * m^{3p}) + \alpha_8 m^{4p} + \alpha_9 m^{5p}. \end{aligned} \tag{2.10}$$

Further, within the general form (2.10) an extensive grid search to find the best values of the power transformation parameters s and p was performed. The type I error rates from the selected best models are presented in Table 4.2. The rates were calculated using validation data with sample sizes of $n_{val} = 45, 65, 85, 100, 250$.

Table 2.1: Performance summary of the correction to the Adjusted Box-Pierce.

Sample size	s	p	AdjBoxPierce	Corrected version
$n = 45$	0.2	0.3	0.04868907	0.05001953
$n = 65$	10.0	1.0	0.05163921	0.05002905
$n = 85$	7.0	2.0	0.05305157	0.05045904
$n = 100$	1.3	1.7	0.05447408	0.05020469
$n = 160$	0.8	0.9	0.05629981	0.04987525
$n = 250$	1.9	0.8	0.05813593	0.05037286

Table 2.2: Summary statistics for selected variables in interval sample size less than 50.

Variable	Estimate	Std.Error	t-value	p-value
n^s	0.425295	0.251604	1.690	0.095008 .
m^p	-1.353900	0.793110	-1.707	0.091837 .
$n^s * m^p$	0.593460	0.396921	1.495	0.138960
$n^{2s} * m^{2p}$	0.149028	0.056476	2.639	0.010065 *
n^{2s}	-0.183531	0.122355	-1.500	0.137706
$n^{3s} * m^{2p}$	-0.070355	0.030893	-2.277	0.025539 *
$n^{3s} * m^{3p}$	0.004419	0.002064	2.141	0.035436 *
m^{4p}	-0.017762	0.004355	-4.079	0.000109 ***
m^{5p}	0.002106	0.000461	4.570	1.83e-05 ***

Table 2.3: Summary statistics for selected variables in finite sample size between 51 and 70.

Variable	Estimate	Std.Error	t-value	p-value
n^s	-2.652e-06	8.296e-07	-3.196	0.00179 **
m^p	1.209e-03	2.984e-04	4.053	9.12e-05 ***
$n^s * m^p$	-2.283e-07	7.347e-08	-3.108	0.00237 **
$n^{2s} * m^{2p}$	-2.068e-12	3.852e-13	-5.369	4.07e-07 ***
n^{2s}	4.910e-10	1.869e-10	2.627	0.00977 **
$n^{3s} * m^{2p}$	4.637e-16	8.877e-17	5.223	7.75e-07 ***
$n^{3s} * m^{3p}$	-1.167e-18	2.439e-19	-4.784	5.05e-06 ***

m^{4p}	6.138e-10	2.856e-10	2.150	0.03364 *
m^{5p}	2.552e-12	1.811e-12	1.409	0.16150

Table 2.4: Summary statistics for selected variables in finite sample size between 71 and 90.

Variable	Estimate	Std.Error	t-value	p-value
n^s	3.214e-17	2.901e-17	1.108	0.269585
m^p	3.833e-06	1.130e-06	3.392	0.000877 ***
$n^s * m^p$	-1.392e-20	3.309e-20	-0.421	0.674609
$n^{2s} * m^{2p}$	-4.627e-36	6.406e-37	-7.224	2.02e-11 ***
n^{2s}	-6.756e-31	6.616e-31	-1.021	0.308740
$n^{3s} * m^{2p}$	9.423e-50	1.523e-50	6.189	5.00e-09 ***
$n^{3s} * m^{3p}$	-1.759e-54	4.077e-55	-4.315	2.80e-05 ***
m^{4p}	2.816e-17	2.774e-18	10.153	$<$ 2e-16 ***

Table 2.5: Summary statistics for selected variables in finite sample size between 91 and 120.

Variable	Estimate	Std.Error	t-value	p-value
n^s	5.169e-06	3.434e-06	1.505	0.133211
m^p	1.266e-05	3.809e-06	3.323	0.000994 ***
$n^s * m^p$	-1.569e-09	9.362e-09	-0.168	0.867045

$n^{2s} * m^{2p}$	-2.021e-13	1.482e-14	-13.641	$<$ 2e-16 ***
n^{2s}	-1.216e-08	7.488e-09	-1.624	0.105408
$n^{3s} * m^{2p}$	3.782e-16	3.539e-17	10.687	$<$ 2e-16 ***
$n^{3s} * m^{3p}$	-4.778e-20	4.874e-21	-9.804	$<$2e-16 ***
m^{4*p}	3.367e-15	1.792e-16	18.793	$<$ 2e-16 ***
m^{5p}	-4.058e-19	3.561e-20	-11.397	$<$ 2e-16 ***

Table 2.6: Summary statistics for selected variables in finite sample size between 121 and 200.

Variable	Estimate	Std.Error	t-value	p-value
n^s	5.966e-05	2.343e-05	2.546	0.01102 *
m^p	8.195e-04	5.830e-05	14.056	$<$ 2e-16 ***
$n^s * m^p$	-1.227e-05	1.336e-06	-9.181	$<$ 2e-16 ***
$n^{2s} * m^{2p}$	-8.989e-09	3.701e-10	-24.290	$<$ 2e-16 ***
n^{2s}	-1.271e-06	3.925e-07	-3.237	0.00124 **
$n^{3s} * m^{2p}$	1.864e-10	5.775e-12	32.280	$<$ 2e-16 ***
$n^{3s} * m^{3p}$	-1.079e-12	2.925e-14	-36.873	$<$ 2e-16 ***
m^{4p}	1.233e-09	8.712e-11	14.147	$<$ 2e-16 ***
m^{5p}	6.308e-12	6.042e-13	10.440	$<$ 2e-16 ***

Table 2.7: Summary statistics for selected variables in finite sample size between 201 and 300.

Variable	Estimate	Std.Error	t-value	p-value
n^s	1.740e-07	5.213e-08	3.338	0.000868 ***
m^p	2.056e-04	5.313e-05	3.870	0.000114 ***
$n^s * m^p$	1.206e-08	5.327e-09	2.263	0.023777 *
$n^{2s} * m^{2p}$	-9.680e-14	6.970e-15	-13.889	$<$ 2e-16 ***
n^{2s}	-1.845e-11	2.884e-12	-6.396	2.22e-10 ***
$n^{3s} * m^{2p}$	5.841e-18	1.928e-19	30.295	$<$ 2e-16 ***
$n^{3s} * m^{3p}$	-5.966e-20	2.469e-21	-24.161	$<$ 2e-16 ***
m^{4p}	-4.111e-09	5.612e-10	-7.326	4.14e-13 ***
m^{5p}	1.660e-10	7.322e-12	22.678	$<$ 2e-16 ***

2.2 Results

Noticeable differences between the patterns of type I error rates across the analyzed sample sizes (40 to 300) were discovered. Therefore, sample-size specific models (0-50, 51-70, 71-90, 91-120, 120-200, 201-300) were constructed to capture the exact pattern for that particular scenario. Table 4.2 displays a condensed form of the comparative study between revised version of Box-Pierce, which to the best of our knowledge is the last version, and the correction that we have brought into the study. For different time series lengths, the corresponding s and p values along with the type I error rates for the adjusted Box-Pierce and those of the corrected version that we designed. It is important to realize that the results from the

implementation of these models show that in all settings, the proposed regression-based correction provided almost perfect type I error rates. In particular, the adjusted type I error rates after the correction to the rejection regions were exactly 0.05 with detailed results.

Tables 2.2, 2.3, 2.4, 2.5, 2.6, and 2.7 show detailed summary from the sample-size specific model fits. These models provide a parametric correction of the type I error rates. Graphical representation of results from the implementation of these models for several scenarios are shown in Figure 2.1.

Form left-to-right-up-to-down the fitting curves with appropriately found models in cases where $(n = 50, 70, 90, 120, 300)$ can be viewed. Empirically, it can be seen that the models that best fit the specific curve in a given data were found.

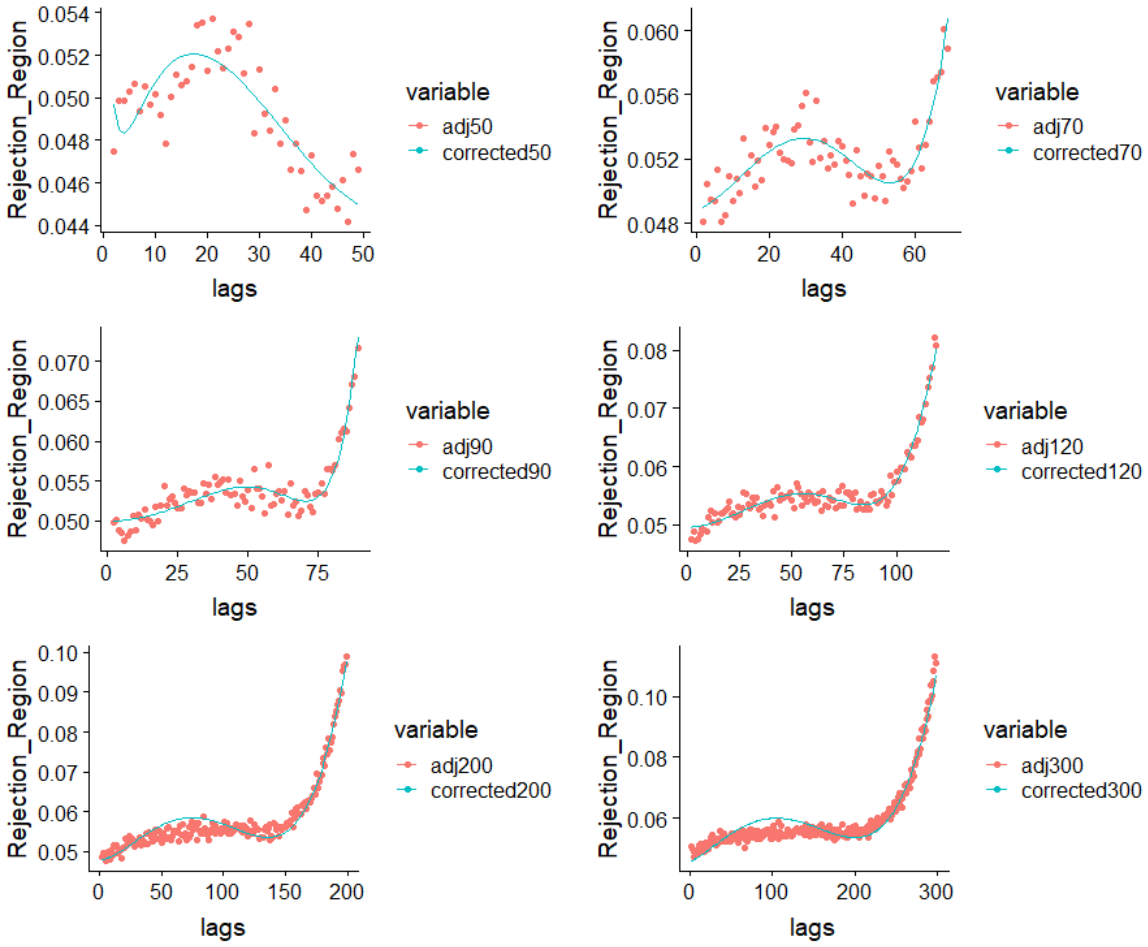


Figure 2.1: Parametric correction to the rejection region

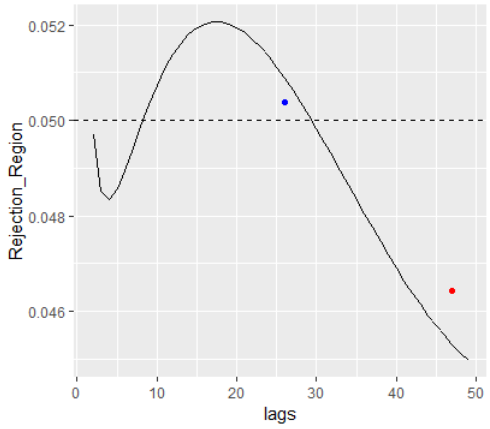


Figure 2.2: Parametric correction to the rejection region

2.3 Data example

An application of our corrected version of the adjusted Box-Pierce test was performed using S&P 500 stock data. We provide instances of both false positive and false negative results obtained by the standard adjusted Box-Pierce test using EQT Corporation stock. This corporation created in 1884 and headquartered in Pittsburg is one of the leading companies extensively devoted to the exploration and transportation of hydrocarbon (Petroleum, natural gas, natural gas liquid). The average daily price of the EQT Corporation was calculated by collecting its opening and closing prices over a period over 8 years (2010-2018). For a window size of 50, numerous false negative and false positive points were found at different lags. In this case, instead of a critical value we have a critical boundary or curve exists. In this setting, the same rejection conditions are the same as in the normal case.

In Figure 2.2, instances of a false positive rejection at lag 26 are shown where the adjusted Box-Pierce test obtains a p-value of 0.0504 but the proposed model correction inflates the rejection region to start at 0.058. The graph also shows a false negative results with p-value of 0.046 at lag 47. However, the proposed correction shrinks the rejection region to start at 0.045.

2.4 Discussion

In this work a new apparoch for correction of adjusted Box-Pierce test recently developed by Kan and Wang [4]. Conceptually, the rejection region correction idea is similar to the ones successfully employed in the work of [16] and [17] to counterbalance the conservativeness of exact homogeneity tests. The provided method combines large scale simulations with subsequent scenario-specific regression modeling that includes complex interaction terms to achieve exceptionally good fit that entails nominal type I error rates for all sample sizes and lags used in the test statistic. The regression models that were constructed depend on the length of the series (n) and the lag order (m). The exponents (s) and (p) of different variables present in the models are treated as hyperparameters in order to control the learning process. To obtain optimal values for those hyperparameters an extensive search through chosen subset values for (s) and (p) was conducted. The simulation study showed that the test outperforms all existing competing goodness-of-fit approaches for sample sizes up to 300.

The merit to the novel correction to the adjusted Box-Pierce proposed in this study is that it allows to find a test with vastly improved type I error rates. This proposed technique of rejection region correction has direct implication on precise decision making by investors and financial institutions. The same technique can be easily extended to larger sample sizes.

2.5 Summary

Building models forces us to translate our beliefs into the language of mathematics and/or computer. More often than not, our believes are erroneous since they are based on assumptions. In this process, we can either make under or over assumptions that could lead to possible misleading us to wrong models with highly destructive consequences. To avoid

eventual undesirable situations, it is always wise to check for the adequacy of our constructed models.

To obtain a dream model, notable mathematical, computational and empirical techniques have been proposed. Some rely on the graphical representation of the estimated residuals and other focus on the plot of the residual autocorrelation and partial autocrrelation at a certain number of lags. However, the first method could be subjective thus deceiving, and the second method only looks at the magnitude and significance of the autocorrelation coefficients at the individual but not jointly.

In trying to overcome the underscored weaknesses of the above mentioned diagnostic of Goodness fit tests, more robust techniques such as Box-Pierce, and Ljung-Box have also been proposed. Despite their success and mathematical soundness, the classical Box-Pierce and Ljung-Box tests for auto-correlation of residuals also present serious flaws, such as severe deviations from nominal type I error rates. As a response, many efforts have been deployed to address this issue by either revising existing tests or designing new techniques. Among all the refined versions of Ljung-Box tests, the Adjusted Box-Pierce demonstrated a superiority by achieving the best results with respect to attaining type I error rates closer to nominal values. Nevertheless, the Adjusted Box-Pierce seems to reject too much.

In this work, we proposed a further correction to the adjusted Box-Pierce test that possesses near perfect type I error rates. The approach is based on an inflation of the rejection region for all sample sizes and lags calculated via a linear model applied to simulated data that encompasses a large range of data scenarios. Our results show that the new approach possesses the best type I error rates of all goodness-of-fit time series statistics.

3 New Risk Factors for Emergency Department Return Visits within 72 Hours for Children with Respiratory Conditions

3.1 Introduction

Chronic respiratory diseases constitute a set of conditions that mainly affect the airways and other parts of the lung. Chronic respiratory diseases have become a complex worldwide epidemiological phenomenon that is highly associated with increased morbidity and mortality[18], [19], [20]. Comparative information about the disease prevalence, visits and returns to emergency departments, financial and death rates show that chronic respiratory diseases have become one of the biggest public health and economic burdens claiming 2.5 million lives and costing $8.1 billion in health care costs in 2019 [21], [22], [23], [24], [25]. Globally, more than one billion people suffer from chronic respiratory diseases and an estimated 4 million die each year [26]. Chronic respiratory infections are reported to be the leading cause of mortality and morbidity in low- or middle-income countries (LMICs) [21], [22], [23], [24], [25] and [26]. As a result, respiratory infections have become an increasingly important part of the global public health efforts and research.

Asthma is another dangerous and prevalent condition with global burden of disability [27], [28], [29]. The Global Asthma Network (GAN), which aims to reduce the suffering from asthma by preventing asthma case with focus on low- and middle-income countries (LMICs), reported that almost 339 million people are affected with Asthma [30]. Annually, asthma is responsible for the death of 489,000 people [21]. With its high rate of mortality and morbidity, asthma is ranked among the top 20 causes of years of life lived with disability. In addition to the death tolls, asthma treatment has high economic burden [31].

Acute lower respiratory infections (ALRIs) kill approximately four million people every year, with higher impact on low and middle income countries [32] placing them among the top three causes of death around the world [30]. For example, in 2015, approximately 2.38 million people lost their lives because of ALRIs [33]. Recently, considerable amount of efforts has been deployed to reduce the death tolls of ALRIs, nevertheless their importance remains underestimated mostly in some regions of the globe [33].

Tuberculosis (TB) has long been viewed as one of the most dreadful preventable infectious diseases. Tuberculosis infects approximately 2 billion people of whom an estimated 2 million of people die yearly [34]. According to WHO, the costs of TB treatment and management represent a disproportionately high burden to patients, their families and communities and governments. For example, each year, tuberculosis accounted for approximately $21 billion including $9.2 billion for treatment and control activities and $12 billion in additional economic costs and lost productivity.

Lung cancer is one of the cancers with the highest mortality rate which claims more lives than any other type of cancers [35]. In 2012, there were 1.8 million new cases and 1.6 million deaths. In 2015, 1.7 million lives were lost due to lung cancer, representing almost 20% of all cancer-related deaths in the world [36].

Sadly, a great number of patients with chronic respiratory conditions are newborn babies and small children. Recent research efforts have reported that chronic respiratory diseases are the conditions related with the highest death rate among children under five years of age. Chronic Obstructive Pulmonary Disease (COPD), though being conventionally accepted to be associated with smoking among quadragenarians or older patients[37] also affects many children. In some studies, it is reported that COPD could take its roots in childhood by living with adult smokers in the same household [38]. Further, it has been reported that there is a strong association between childhood death from COPD and poverty [39].

Asthma is the most common, but non-communicable chronic disease that not only impacts the quality of life of children, but also significantly contributes to childhood mortality and morbidity worldwide [40], [41] with a death rate ranging between 0.0 to 7.0 per 1,000 [24]. On a global scale, 14% of children are affected with asthma and as high as one-third of US children population [42].

Presently, there is no known cure for asthma. However, the disease can be managed with adequate prevention and treatment therapy. According to recent Center for Disease Control (CDC) data, 1 in 13 people have asthma resulting in ∼25 million Americans having the disease and sadly, children account for ∼8% of the patients. Asthma has been on a steady increase in the last three decades affecting all ages, gender, and ethnicity. Asthma is the leading chronic disease in children, and it is more common in boys than in girls. Strikingly, it is the top reason for school absenteeism among school children. In 2013, ∼13 million missed school days were attributed to asthma. In 2015 and 2016 CDC data, it was observed that 48% of children ages 18 and under who had the disease reported having at least one asthma attack in the previous year. Likewise, ∼50% of children under the age of 5 with asthma had an episode.

The financial implications of asthma are enormous with an annual economic cost in the worth of $80 billion between 2008 and 2013 [31]. The annual per-person incremental medical cost of asthma was approximately ∼$3,200 in 2015 [31].

Globally, the prevalence of asthma in children has been on the rise [43], [44]. Burr et al., between two different studies, outlined an increase of 6.5% in the prevalence [45]. A similar trend in the prevalence of asthma was reported by Burney and Aberdeen in [46], [47]. With an increasing prevalence, asthma has progressively become the most frequent cause of hospitalization among children [48]. In theory, the prevalence of asthma symptoms is assumed to attain higher levels in developed or high-income countries (HICs), however

LMICs have recently displayed alarming prevalence rates. Recent studies showed that more severe childhood asthma cases are seen in LCMIs with over 80% of asthma-related deaths occur in LMICs [49], [50].

Tuberculosis (TB) is another respiratory infection causing a pronounced increase in morbidity and mortality in children around the world [51], [52], [53] particularly among those from geographic locations with high incidence and prevalence of HIV [49], [54]. Due to poor recording and reporting of childhood TB cases, lack of resources, and limited amount of pediatric surveillance data, quantifying and estimating accurately the global pediatric burden of TB has been subject of great debates [55], [56], [57]. Nevertheless, a staggering number of pediatric pulmonary tuberculosis cases have been encountered. For example, in 1989 the World Health Organization (WHO) claimed that 450,000 deaths in children under 15 years of age occur almost every year [58]. In 2000, an estimated 8.3 million new cases of tuberculosis and 1.8 million of deaths were reported [56], [59]. In 2014, WHO indicated that approximately one million new cases occurred among children, of whom 136,000 died [60]. Though considerable efforts have been recently deployed to control the prevalence and incidence of TB cases, tuberculosis remains a serious public health challenge.

The exacerbation of chronic pediatric respiratory conditions is generally associated with unscheduled returns to the emergency department (ED) within 72 hours. By return visits to the ED, we refer to the definition given by De Sales et al. as the return of a patient to the ED because of the initial complaint within 72 hours from being discharged [61]. Though unplanned return visits to the ED are generally related to the progression of the illness [62], [63], many other factors and scenarios including the quality of the health systems that take patients in, medical errors, persistence of the parents for extra-care may also be involved [64], [65]. Another important concern is that ED 72-hour revisits impose burdens not only on patients and their families, but also to insurance providers. Unscheduled readmissions may lead to overcrowding of medical facilities and incur financial burdens on hospitals [66],

[67]. Return to ED rates have been used as a metric to assess the quality care provided to patients, where higher rates are widely used to designate inappropriate treatments or eventual medical errors [62]. Today, the rate of return visits to the ED is the metric of choice for measuring care quality in the ED. In recent literature, a generally accepted rate of return visit to the ED is estimated to be less than 1% [61]. According to other researchers, return visit rate ranges between 2.5% and 5.2% for emergency departments [66], [63], [68]. In contrast, [69] indicated that site-specific 72-hour return visit rates ranged from 1.1% to 15.2%. A great number of the previous studies related to the association between chronic respiratory conditions and the rate of ED return reported some factors such as young age, health insurance coverage, higher acuity to be the main causes for returning visits to the ED [70], [66]. Other studies have reported that the greater rate of returning to the ED is significantly associated to young age [66], [63].

Little research has been conducted to investigate new risk factors for return ED visits for patients with existing respiratory conditions. Nevertheless, [71] reported that antimicrobial prescription for upper respiratory infections among patients covered by Medicaid has decreased, and that there is no association between the prescription and the decrease in subsequent return visits. [63] studied the frequency of pediatric 72-hour return visits to the ED between 2001 and 2007. A significant increase in the return visit rate was noticed between 2001 and 2007. The authors also found factors such as age, arrival time to the ED, recent discharge from the hospital, and some geographical regions of the US to be strongly associated with return visits to the ED.

In this study, we provided comprehensive analyses of return visits in the pediatric ED settings among children with chronic respiratory conditions. We identified all statistically significant respiratory conditions that are predictive for return ED visits among children that have been discharged within 72 hours. We estimated the magnitude and directions of the effects sizes, allowing hospitals, healthcare facilities and public health institutions to design and adopt

a more accurate and advantageous regulation for handling high risk patients. Lastly, we implemented several machine learning algorithms to find the best predictive model for ED readmission within 72 hours for children with existing pulmonary conditions.

3.2 Methods

We conducted this multicentered epidemiological study using data queried from the large Health Facts database. This database is a repository of de-identified medical data from 650 hospitals centers throughout the United States containing complete details of all patient visits since 2015. The existing data in the database are obtained from electronic medical records, which are also provided by contributing hospitals and organizations. These records can include encounter data (emergency, outpatient, and inpatient), provider specialty, demographics (age, sex, and race), diagnoses and in-hospital procedures documented by ICD-9-CM codes, laboratory data, pharmacy data, in-hospital mortality, and hospital characteristics [72]. While a range of hospitals and other medically related entities from different horizons around the country collect the data, Cerner Corporation is the principal representation that mainly captures, centralizes, and stores the data. As of 2018, Health Facts brings together information from 90 health centers and 650 facilities across the states. In this study, we however decided to use the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) to determine patients treated from respiratory conditions not associated with cancer and as captured by ICD-10-CM codes J00-J99.

We conducted a retrospective case-control study of patients under 18 years of age nested in the larger cohort of all patients. The selected patients were admitted in 166 emergency departments. The approval to conduct this was given by the CHOC Children's Hospital Institutional Review Board (IRB 180857). The used deidentified dataset contains the admission data of approximately 1.7 million patients that were admitted in pediatric emergency

departments (EDs). The available demographic variables were race (Caucasian, Hispanic, African American, and Black, Asian Pacific Islander, Native American, unknown), age and sex. The demographic variable age was divided into three categories: (0, 5], (5,12], and (12,18]. Patients whose sex information was not available were excluded from the study. The type of insurance was used to approximate the individual patient's socioeconomic status. The patients involved in this study use one of the insurance types (commercial, Medicare Medicaid, 2 other government insurance types, self-pay, and others). Since chronic respiratory conditions can result in unplanned return to the ED and higher hospital utilization, related variables were added. The key reason for this is that hospital utilization generally relates to the efficiency of the care-quality received by patients. As such, the hospital utilization variables used as explanatory variables are Previous ED visit, Ispediatric, Has History of Return Visit and number of medicines taken by the patient, and length of stay (LOS). The patient's length of stay was also divided into four categories: (0, 4], (4,8], (8,16], (16,24]. Some of the most common respiratory conditions coded as Acute Nasopharyngitis[common cold] (J00), Acute Sinusitis(J01), Acute Pharyngitis (J02), Acute Tonsillitis (J03), Acute Laryngitis (J04), Acute Obstructive Laryngitis and Epiglottitis (J05), Acute Upper Respiratory Infections of Multiple and Unspecified Sites (J06), Influenza due to certainidentified influentza viruses (J09-J11), Viral Pneumonia, not elsewhere classified (J12), Bacteria and Other Pneumonia (J13-J18), Acute Bronchitis(J20), Acute bronchiolitis (J21), Vasomotor and Allergic Rhinitis (J30), Chronic Rhinitis, Nasopharyngitis and Pharyngitis (J31), Chronic Sinusitis (J32), Chronic disease of tonsils amd adenoids (J35), Nasal Polyp Other Unspecified Disorders Nose Nasal Sinuses (J33-J34), Peritonsillar Abscess (J36), Other diseases of Upper Respiratory tract (J39), Bronchitis, not specified as acute or chronic (J40), Simple and Mucopurulent chronic bronchitis (J41), Unspecified Chronic Bronchitis (J42), Emphysema (J43), Asthma (J45), Acute Respiratory Distress Syndrome (J80), Suppurative and Nectrotic of Lower Respiratory Tract (J85-J86), Pleural effusion, plague, and other

pleural conditions (J90-J92, J94), Intraoperative and Postpreocdural complications and Disorders not classified elsewhere (J95), and Other diseases of the respiratory system (J98-J99) were included in the study. Table 3.2 displays the prevalence rate of the concerned respiratory conditions. addition, Surgical procedures during the index ED visit were not included in the study, but they were also classified into auditory, cardiovascular, digestive, integumentary, musculoskeletal, and urinary/reproductive system surgery using the current procedural terminology code version 4 (CPT-4). Furthermore, we estimated the 90th percentile of the number of medications administered during the ED visits and categorized patients into 2 groups: patients who received greater than the 90th percentile and those that did not. Lastly, we excluded data from all ED facilities that have had less than 100 return visits as they corresponded to facilities with disproportionately small number of encounters and may be a result of data entry error and noise.

As seasonal variation is known to be also responsible for clinical discomfort to patients with chronic conditions and an increase of the hospitalization rate and that of return to ED, we decided to categorize the variable that describes the month at which the patients were readmitted in ED into 4 categories. Winter (December 1st- February 28th or 29th in leap year) was mapped to 0, Spring season (March 1st – May 31st) to 1, summer (June 1st – August 31st) to 2, and Autumn (September 1st – November 30th) to 3. Patients were excluded if they have respiratory condition occurrence rate less than 1,000 and several medications ($>$ 10). The primary reason for this is to make sure that we include only facilities that have seen large number of patients with respiratory conditions. The secondary reason for this selection technique is also exclude noise in the data in relation to ED facilities. Subjects who also spent more than 24 hours in ED were also dropped in the study.

In the study, the multicollinearity in the data was assessed by estimating the generalized variance inflation factor (GVIF) of each of the variables. GVIF is a statistical tool that quantifies the degree of correlation between the predictors present in a given model. One

mission of GVIF is to identify variables that are highly correlated with each other. Another mission of GVIF is to be able to assess the contribution of involved variables in the model. Addressing the degree to which variables are correlated (multicollinearity) improves the measurement of association between a variable within a model and the outcome it is predicting. This decision was made, a priori, to exclude all variables with GVIF greater than 4 – a rule of thumb threshold based on previous studies [73], [74], [75].

The data used in this multicentered epidemiologic study were prepossessed HealthDataLab using Center Corporation – an elastic parallel distributed high-performance cloud computing platform running on Apache Spark Since our data are clustered by Hospital ID and Patient ID, a mixed random effects logistic regression model was deployed to conduct a multivariate analysis. Thus, a random intercept model was built to model the return to the ED within 72 hours, a binary outcome variable. The model building and statistical analysis were preformed using R statistical Software. The machine learning models were constructed using Python Computing Programming Language

We implemented 15 high-accuracy machine learning classiication models such as Decision Tree (DT), K-Nearest Neighbors (KNN), Gaussian Naïve Bayesian (GNB), Multinomial Naïve Bayesian (MNB), Complement Naïve Bayesian (CNB), Multinomial Logistic Regression (MLR), Multi-Layer Perceptron (MLP), Ridge Regression Classifier (RRC), Linear Classifiers with Stochastic Gradient Descent (LCSGD), Passive Aggressive Classifier (PAC), Linear SVC (SVC), Random Forest (RF), Extremely Randomized Trees (ERT), Gradient Boosting Tree (GBT), and Extreme Gradient Boosting Tree (EGBT) on the task of classifying subjects as patients that have been readmitted, or not readmitted 72 hours after discharge. Using grid-search, hyper-parameter tuning optimization across different models was carried out. The optimal values of the hyperparameters were selected based on the AUC over 10-fold validation datasets. The implementation of the adopted machine learning algorithms and the hyper-parameter tuning were done in python and while using scikit-learn.

Table 3.1 displays the names of the machine learning algorithms involved in this study, the names of the considered hyperparameters and the values and options they assume. The choice of these machine learning algorithms, associated hyperparameters and their values was inspired by a study recently done by Zhen and colleagues (Zheng et al., 2020).

Table 3.1: Hyperparameters Table

Model Name	Hyperparameter Name	Hyperparameter Options
DT	criterion	'gini','entropy'
	splitter	'best','random'
	max_features	'auto','sqrt','log2',None
KNN	n_neighbors	15 31
	weights	'uniform','distance'
	algorithm	'ball_tree','kd_tree'
GNB	var_smoothing	$10^{-7\sim-12}$
MNB	alpha	0, 0.1, 0.5, 0.8,1
CNB	alpha	0, 0.1, 0.5, 0.8,1
MLR	solver	'newton-cg','lbfgs','saga','sag'
RRC	alpha	1e-3, 1e-2, 1e-1, 1
	solver	'svd', 'cholesky', 'lsqr', 'sparse_cg', 'sag', 'saga'
LCSGD	loss	'hinge','log','modified_huber','squared_hinge','perceptron'
	alpha	1e-3, 1e-2, 1e-1, 1
	learning_rate	'constant','optimal','invscaling','adaptive'
	eta0	0.01, 0.001, 0.0001
PAC	C	0.001, 0.01, 0.1,1
	loss	'hinge','squared_hinge'
SVC	loss	'hinge','squared_hinge'
	C	0.001, 0.01, 0.1, 1
RF	n_estimators	300, 500, 800
	criterion	'gini', 'entropy'
	bootstrap	True, False
	max_features	'auto', 'sqrt', 'log2',None
ERT	n_estimators	300,500,800

	criterion	'gini', 'entropy'
	bootstrap	True, False
	max_features	'auto', 'sqrt', 'log2',None
GBT	loss	deviance, exponential
	learning_rate	0.1 , 0.01, 0.001, 0.1
	subsample	0.1, 0.5, 0.9
	n_estimators	300, 500, 800
	max_features	'auto', 'sqrt', 'log2', None
EGBT	tree_method	'auto', 'exact', 'approx', 'hist'
	grow_policy	'depthwise', 'lossguide'
	n_estimators	300, 500, 800
	learning_rate	0.001, 0.01
	max_depth	10, 15, 20, 50, 100
MLP	hidden_layer_size	(50,50,50), (50,100,50), (100,)
	activation	'tanh', 'relu'
	solver	'sgd', 'adam'
	alpha	0.0001, 0.05
	learning_rate	'constant', 'adaptive'

Figure 3.1: ED return proportions with respect to patient's age

3.3 Results

To make inference about the whole enrolled patients and identify statistical and causal associations between variables that are present in the data at the time of the study and the ED return within 72-hours after discharge to home, univariate and multivariate analyses are then performed. As such, refers to Table 3.2, and Table 3.3 for results generated from the univariate study, and to Table 3.4, and Table 3.5 for outputs obtained from the multivariate counterpart.

Patients and ED return Rates

In this study, the total sample size was 1,513,333, where the subjects were distributed among 166 hospitals across the nation. A total number of 48,828 (3.23%) returned to ED within 72 hours versus 1,464,505 (96.77%) who did not return to ED at least in the next 72 hours. Thus, ED readmission is a relative event that induces imbalanced data. Per our modeling objective, we partitioned the obtained dataset into training and validation sets in 75-25%

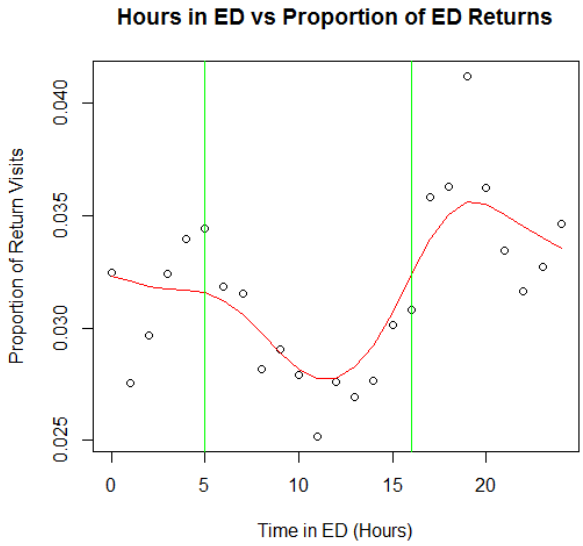


Figure 3.2: Return proportions with respect to the time spent by patients in ED

ratio respectively. The train set was used for learning purposes, while the remainder of the data was used to validate the models. Exploratory analysis and descriptive statistics of the demographic and clinical characteristics of the involved subjects present in the training set are displayed in Table 3.2, and Table 3.3. Among all the eligible patients in the training set, 46.51% were female while 53.49% were male. Of these patients 42.36% were Caucasian, 31.22% were African American, 7% were Hispanic, 1.24% were Asian Pacific Islander, 2.34% were Native American and 16% were of non-identified ethnic or racial group(s). Approximately, 57.60% of the eligible subjects were on Medicare or Medicaid versus 23.32% on Commercial Health Insurance, and 10.77% in other types of insurance. Only 6.03% of the children were either covered by their family and 2.26% of the patients were under other governmental insurance. The majority (57.67%) of the patients were 5 years of age or less, against 29% aged between 5 to 12 and 14% of the encounters older than 12 years of age. Above all these, the information in the training set indicates that the rate of return to the emergency department within 72 hours after being discharged was only of 0.032.

Demographic and Socio-economic Characteristics vs ED Returns

From our statistical analysis, we observed that the readmission rates in male patients is higher than in female patients as shown in Table 3.2. Strikingly, we found the existence of a significant ($P< 0.001$) association between other demographic characteristics such as participant's age, length of stay, race/ethnicity, and their medical insurance and ED return rates. Participants less than 5 years old were the largest number of participants to return in ED. Figure 3.1, for example, displays the proportion ED returns visit with respect to patient's age. Such a proportion follows a quasi-parabolic shape where the maximum number of unscheduled returns to ED is more pronounced with children that are less than 5 years. The absolute minimum of such a proportion is achieved among children that are 5 to 10 years. Though readmission rate to the ED was higher among children older than 5 and less than 12 years of age, it was not as common as with children of less than 5 years of age. The remaining demographic/Payer/Resourceutilization variables have significant association with pediatric patients revisiting the ED. It can then be concluded that demographic variables, proxies for socio-economics status and those for hospital resources are eminent risks factors to the ED Returns(Figure 3.2).

From our multivariate analysis, we observed that the time spent by patients in the ED during the last six (6) months is associated with higher odds of returns to the ED. Interestingly, we discovered that the longer the length of stay, the higher the odds of returns to the emergency department became. Patients that spent 16 to 24 hours in the ED have 30.4% increase in odds of return, followed by the ones that spent 8 to 16, and 4 to 8, with 15 and 13% odds of return. Next to the length of stay, proxies of socio-economic statuses come in term of being associated with the risk of return to ED. The model reveals that besides patients, regardless of the type of insurance they own, have an increase in the odds of an unscheduled readmission to the ED returns. Patients that have insurance of type "others" have the highest increase in odds (20%), followed by those who have other governmental insurance, Medicare or Medicaid

with, approximately, 14-, 13-, 4% respectively of likelihood to unexpectedly return to the ED. Also, we observed that patients that were of ages 5 to 12 had higher drop (22%) in the odds of a return visit than patients that are older than 12 years of age who present have a 19% drop in odds of a return visit. Furthermore, we examined the importance of race and ethnicity in the ED. Strikingly, we found that patients of with African American descent had a 6% decrease in the odds of a return visit. The remaining ethnic group did not display any statistical significance. Patient's past hospitalization, previous ED visits and history of return visits are highly associated with the risk of a return visit. In addition, they all have an increase in the odds of return visit ranging from 16% to 124%. The free-standing pediatric ED variable appears to not be statistically significant.

We also found that the number of medications administrated the patients is statistically and significantly associated with the odds of a return visit in the ED within 72 hours. In multivariate analysis, we found that patients that are taking more than 3 medications have almost 40% increase in the odds.

Seasonal effects vs ED Returns

From our examination of seasonal effects and ED returns, we realized that seasonal effects significantly impact the rates of returning to the ED (Table 3.2). For example, the ED return visit rate varied by periods of the year. The ED return visit rate is higher in Winter (32%) followed by ∼19, ∼50 and ∼19% decrease in Spring, Summer and Fall semesters respectively. However, there is a decrease in the ED return visit rate (∼38%) from Spring to Summer against an increase of almost 63% between Summer and Fall. Overall, seasonal effects should be taken as a major cause for patients under 18 to return to the ED. In the multivariate analysis, we can see a reduction in odds in Spring and Fall (1.2 and 7.2%, respectively) versus an increase in odds of ED return visit in Summer. Noteworthily, we found that Fall semester significant risk factor that contributes to the ED return visits.

Diagnoses vs ED Returns

In the diagnosis and ED return examination, we found that the rate of return to the ED is significantly ($P < 0.001$) affected by some types of diagnoses (Table 3.2). For example, besides Mental and Behavioral disorders (F01-F99), Diseases of the nervous system (G00-G99), Diseases of the musculoskeletal system connective tissue (M00-M99), Pregnancy/Childbirth/Puerperium (O00-O9A), some conditions related to the perinatal period(P00-P96) and Congenital and Chromosomal abnormalities(Q00-Q99), the remaining diagnoses constitute statistically significant ($P < 0.001$) risk factors to the ED return visits. The rate of return visits was higher with patients diagnosed with Diseases of the eye and Adnexa (H00-H59), infectious/Parasitic diseases (A00-B99) and diseases of the digestive system(K00-K95), 13, 9, and 8%. In multivariate analysis, we found that apart from Congenital chromosomal abnormalities (Q00-Q99), Injury/Poisoning/Consequences of External Causes (S00-T88), and Diseases of the eye and Adnexa (H00-H59) and that have 12, 19 and 20% decrease in odds of return visits, the remaining diagnoses have an increase in odds(between 6 and 71%) of return visits. Of these diagnoses known as risk factors for patients to return to emergency department, only infectious/Parasitic diseases (A00-B99), Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism (D50-D89), Diseases of the eye and adnexa (H00-H59), Diseases of the digestive system (K00-K95), and Diseases of the genitourinary system (N00-N99) are statistically significant.

Respiratory Conditions vs ED Returns

In our respiratory conditions and ED returns analysis, we observed that the relationship between respiratory conditions and rate of ED returns is also outlined (Table 3.3). Respiratory conditions associated with significant ED returns are : Acute sinusitis (J01), Acute pharyngitis (J02), Acute tonsillitis (J03), Acute obstructive laryngitis and epiglottitis (J05), Acute upper respiratory infections of multiple sites (J06), Influenza (J09-J11), Viral pneumonia (J12),

Bacteria and other pneumonia (J13-J18), Acute Bronchitis (J20), Acute Bronchiolitis (J21), Vasomotor and allergic rhinitis (J30), Chronic rhinitis/nasopharyngitis/pharyngitis, Chronic sinusitis (J32), Nasal polyp/other nose or nasal infections (J33-34), Peritonsillar abscess (J36), Other chronic lower respiratory diseases (J40-J43) and Intraoperative/postprocedural complications/disorders of the respiratory system (J95). Patients diagnosed with Acute upper respiratory infections of multiple sites (J06) and Acute pharyngitis(J02) had the highest proportion of returns (37% and 18%, respectively). Alongside these conditions are aligned Acute bronchiolitis (9%), Nasal poly/other nose or nasal infections (8%), and Bacteria and other pneumonia (7%) as conditions with high ED revisit rates. In multivariate analyses, patients who suffered from Peritonsillar Abscess (J36), Viral pneumonia, not elsewhere classified(J12), Acute bronchiolitis (J21), Bacteria and other pneumonia (J13-J18) and respiratory conditions (J90-J92, J94) have higher increase odds (127%, 44%, 39%, 27.3% and 16% respectively) of a return visit.

Patients diagnosed with Acute pharyngitis, Asthma (J45, though not being statistically significant in univariate case), Acute upper respiratory infections of multiple sites, Acute nasopharyngitis/common cold have reduced 5, 8, 9, and 10% decrease in the odds of a return visit. It was also found that certain respiratory conditions such as Acute tonsillitis (J03), resp (J4044COPD1), Chronic sinusitis (J32) and Acute bronchitis are associated with reduced odds (10, 11, 14 and 18%, respectively) of a return visit. Conditions like Acute sinusitis (J01), diseases of upper respiratory (J39), and Chronic diseases of tonsils and adenoids (J35) have also reduced odds (23, 26, and 28%) of a return visit.

Comorbidities Surgical Procedures vs Returns

As of the association of the simultaneous presence of other chronic diseases or conditions and the rate of returns, Table 3.2 indicates that the most common causes for unscheduled returns to ED are Cardiovascular surgery (CPTA:69000-69979), and Urinary/Reproductive system

surgery (CPT4:50010-58999). They constitute serious risk factors of returning to emergency department within 72 hours, with rates of returns 48 and 23%, respectively. Though the readmission rates among patients that went through Integumentary surgery (CPT4:10030-19499), Musculoskeletal surgery (CPT4:20100-29999), Auditory surgery (CPT4:69000 -69979), and Digestive surgery (CPT4:40490-49999) have higher ED rates of returns (15, 14, 13, and 12% respectively), they remain variables that are not statically significant. In the multivariate analysis setting, besides Digestive surgery (CPT4:40490-49999), which is not statistically significant, the remaining surgical procedures were associated with very high risk of return visit. With Patients that underwent cardiovascular, Integumentary and Urinary/Reproductive system surgery (CPT4:50010-58999), there is at least 49% and at most 80% increase in odds of return visits.

Table 3.6 exhibits the average AUC of the 10-fold cross validation testing for the 15 machine learning models chosen in this study. The best performance was achieved by the Extreme Gradient Boosting model over all the considered tested models with AUC of 0.645. Another observation is the poor over all performances of the considered machine learning models. This can be presumably due the inability of machine learning models to effectively predict return visits given the current data under investigation. Another reason for these comparatively poor performances could be that return visits may be influenced by factors that are not clinical in nature but driven by behavioral patterns of the families with these patients.

3.4 Discussion

Unexpected returns to ED within 72 hours after being discharged generate consequential economic and social damages. Therefore, identifying the associated preeminent risk factors that could lead to such undesirable situations has been increasingly growing to be a focal point for parents, medical staff and hospital management as it can, at least, lead to better

further decisions.

The present study found that despite that male patients have a higher chance of revisiting ED than female patients; gender is not associated with higher risk for ED return within 72 hours after being discharged. This is consistent with what was found by [19], [76], [77]. Our model indicates that patient's age is proportional to the drop rate in the odds of a return visit to ED. This tells us that younger patients are more likely to revisit the emergency department. A similar result was reported by [62]. Such a finding is presumably due to the fact that when children are younger, they are expected to have have a weaker immune system.

Interestingly, our model indicates that the duration of a single episode of hospitalization is most likely to be a significant risk factor for ED readmission. Patients that have spent a longer time have a higher chance of revisiting ED within 72 hours after discharge. This result could be related to the fact that the time spent in the medical facility could be a good indicator of unwell the patients could be. Regarding race, African American patients are more likely to return to the ED within 72 hours more than Caucasian patients. In terms of insurance type, patients on Medicare/Medicaid or Other governmental insurance or other types of coverage are more likely to returns to the ED. Our model also indicates that previous ED visits were a very strong predictor of at-risk patients. In fact, the number of previous ED visits was found to be proportional to the risk of returning to the ED. Precisely, patients that have previously visited ED the most are at higher risk of making another return.

In terms of chronic respiratory conditions, we found that patients suffering from complications coded with J01, J06, J09-J11, J13-J18, J20, J21, J31, J36, and J45 are more inclined to return the the ED within 72 hours after being discharged. Identifying patients that display these medical complications is of great importance as it can help clinicians to : (1) know which patients need extra care or special attention, (2) make better orientation decisions,

(3) determine which patients will need post-ED support, (4) determine what type of support should be provided to the concerned patients, and (5) educate the patients risks that could exacerbate their conditions. Nevertheless, it shall be noted that most of the respiratory diagnoses are unlikely to result in an ED revisit.

In this study, we surprisingly found that the presence of some comorbidities was strongly associated higher chance to unexpectedly return to ED with 72 hours after discharge. We noticed found that patients that are suffering from Intraoperative and postprocedural complications and disordered (J95), Cardiocascular surgery (CPT4:33010-19499), Integumentary surgery (CPT4:10030-19499), Unirary/Reproductive system surgery(CPT4:50010-58999) were inclined to revisit the emergency department unexpectedly.

Arguably, building models, regardless of their nature, involves caveats and limitations that are generally related to lack of extra information (i.e data), under or over assumption. Similarly, this study suffers from a variety of weaknesses. Firstly, the data used for this investigation is administrative data. This insinuates the presence of eventual coding errors during the period of assigning diagnosis codes to patients. Secondly, our data set is highly unbalanced in favor of not returning to the ED within 72 hours after being discharged. Thirdly, it should be known that despite the broad geographical distribution of the enrolled EDs across the country, it will be risky to consider the involved population in this study could be taken as a viable sample representative of all the children undergoing chronic respiratory conditions. Therefore, the study findings should not be generalized to all the children suffering from respiratory conditions. Fourthly, the data set that we use for modeling purposes does not tell us exactly how the patients were enrolled in this study. Plus, we do not have any information about patients' care during their previous visits. Knowing this may be help orient the efforts during their next appearance in the emergency department. Fifthly, the cohort of patients in this study was based on patients discharged home from the ED. This implies that we missed patients admitted to the hospital through the ED and return to the ED after

discharge from the hospital. Such a limitation is, however, one of design and related to the question being asked. In our case, we are concerned in elucidating the difference between patients who return to the ED and those who do not among patients discharged home from the ED (and deemed not requiring hospitalization).

Besides the information about the type of insurance used by the patients, no other socioeconomic status is available. For example, the level of education could be used to investigate the degree of awareness of parents with respect to the utilization of ED resources.

Also, this study suffers from the way the study was designed. For example, our mixed random effect model enables determine which variables could be seen as major risk factors of return to the emergency department. However, it does not inform us about which ED return was avoidable. Another limitation that could come from the study design is that some of the return visits that were captured may have been unavoidable which we were not able to capture/establish in this study.

Despite the above highlighted caveats and limitations, the contributions and implications of this study are noteworthy. This study may help to (1) inform the patients about causes and risk factors that could lead to a potential exacerbation of their health conditions; (2) educate parents on the urgency of pediatric conditions; (3) provide EDs with what is necessary to determine which interventions might be needed; (4) measure the effectiveness of the services provided by EDs; (3) proffer ways to improve the quality of care of patients at risk; and (4) construct very solid future strategies that could help drastically reduce potential avoidable ED return visits.

3.5 Conclusion

This multicenter study of pediatric return visits within 72 hours among patients with respiratory conditions across 166 emergency departments established that revisits within such short period of time may be driven more by non-respiratory comorbidities than the underlying respiratory conditions. This pattern may differ as the window is increased to a value greater than 72 hours. However, peritonsillar abscess, pneumonia, and bronchiolitis exposes pediatric patients to higher odds of return visit. These findings indicate that ED providers should pay closer attention to the respiratory risk factors as well as the comorbid conditions that patients with respiratory conditions may present with. Corresponding improvements in the quality of care and in the education of patients and their families may result in reductions in return visits to the ED.

3.6 Summary

Understanding and dealing effectively with complex processes, such as emergency department (ED) return visits, in order to better predict and efficiently minimize associated risks have been the subject of virulent debates and the ground of countless arguments. In the past few decades, academics, health management experts and insurance companies have proposed a forest of techniques for reaching a better destination with respect to with ED return visits. Among other approaches, predicting patients that are more likelihood to have ED return visits has become one of the principal concerns. In this light and in the spirit of taking part of this concert, we intended to identify respiratory conditions and associated comorbidities most likely to result in a return visit among children discharged home from the ED. For this, special attention was given to univariate and multivariate analyses and machine learning consideration. The univariate study was performed in analyzing the summary statistics of

the eligible population. The multivariate analysis was conducted in using a nested mixed effects model, with the aim of modeling the return to ED within 72 hours such that the ED facilities are random intercepts and patients nested within the facilities. The machine learning consideration was executed by hyper-tuning 15 classifiers.

It resulted from the study that unexpected return visits to the ED among children undergoing respiratory is statistically and significantly associated with conditions such as Acute bronchiolitis [odds ratio and 95% confidence interval: 1.39 (1.35, 1.44)], pneumonia [1.22 (1.18, 1.27)], and asthma [1.07 (1.04, 1.10)]. In addition, we found that over 80% of non-respiratory comorbid classes of diseases are associated with increased risk of return visits.

Furthermore, we found that machine learning models are not imperatively suited for predicting return visits as these could be influenced by factors that are related to behavioral patterns of the families with the patients.

Table 3.2: Univariate Summary Statistics (USS)

Variable	Levels	No return visit N (%)	Had return visit N (%)	Chi-squared p value
Resource Utilization, Medications, and Season				
ED length of stay (hours)	[0,4)	934976 (85.15)	30572 (83.21)	< 0.001
	[4,8)	106997 (9.74)	4215 (11.47)	
	[8,16)	41631 (3.79)	1394 (3.79)	
	[16,24)	14428 (1.31)	558 (1.52)	
Previous ED Visit	0	674715 (61.45)	17544 (47.75)	< 0.001
	1	252210 (22.97)	9477 (25.80)	
	2	99013 (9.02)	4567 (12.43)	
	3 or more	72094 (6.57)	5151 (14.02)	
Previous hospitalizations (prior 6 months)	No	1047164 (95.37)	34104 (92.83)	< 0.001
	Yes	50868 (4.63)	2635 (7.17)	
Has History Of Return Visit	No	1038850 (94.61)	32619 (88.79)	< 0.001
	Yes	59182 (5.39)	4120 (11.21)	
Free-standing pediatric ED	No	679870 (61.92)	21307 (58.00)	< 0.001
	Yes	418162 (38.08)	15432 (42.00)	
Number of medications greater than 90th percentile	No	1060102 (96.55)	34853 (94.87)	< 0.001

	Yes	37930 (3.45)	1886 (5.13)	
Season	Winter	350208 (31.89)	12114 (32.97)	< 0.001
	Spring	284611 (25.92)	9628 (26.21)	
	Summer	177002 (16.12)	6081 (16.55)	
	Fall	286211 (26.07)	8916 (24.27)	
Comorbid Diagnoses				
Certain infectious and parasitic diseases (A00-B99)	No	1008346 (91.83)	33356 (90.79)	< 0.001
	Yes	89686 (8.17)	3383 (9.21)	
Neoplasms (C00-D49)	No	1096685 (99.88)	36653 (99.77)	< 0.001
	Yes	1347 (0.12)	86 (0.23)	
Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism (D50-D89)	No	1091372 (99.39)	36275 (98.74)	< 0.001
	Yes	6660 (0.61)	464 (1.26)	
Endocrine, nutritional and metabolic diseases (E00-E89)	No	1083629 (98.69)	36060 (98.15)	< 0.001
	Yes	14403 (1.31)	679 (1.85)	
Mental, Behavioral and Neurodevelopmental disorders (F01-F99)	No	1083399 (98.67)	36171 (98.45)	< 0.001
	Yes	14633 (1.33)	568 (1.55)	
Diseases of the nervous system (G00-G99)	No	1069382 (97.39)	35721 (97.23)	0.058
	Yes	28650 (2.61)	1018 (2.77)	
Diseases of the eye and adnexa (H00-H59)	No	933720 (85.04)	31914 (86.87)	< 0.001
	Yes	164312 (14.96)	4825 (13.13)	

Diseases of the circulatory system (I00-I99)	No	1091760 (99.43)	36416 (99.12)	< 0.001
	Yes	6272 (0.57)	323 (0.88)	
Diseases of the digestive system (K00-K95)	No	1055944 (96.17)	34980 (95.21)	< 0.001
	Yes	42088 (3.83)	1759 (4.79)	
Diseases of the skin and subcutaneous tissue (L00-L99)	No	1069817 (97.43)	35611 (96.93)	< 0.001
	Yes	28215 (2.57)	1128 (3.07)	
Diseases of the musculoskeletal system and connective tissue (M00-M99)	No	1077523 (98.13)	36100 (98.26)	0.076
	Yes	20509 (1.87)	639 (1.74)	
Diseases of the genitourinary system (N00-N99)	No	1082372 (98.57)	36051 (98.13)	< 0.001
	Yes	15660 (1.43)	688 (1.87)	
Pregnancy, childbirth and the puerperium (O00-O9A)	No	1097151 (99.92)	36689 (99.86)	< 0.001
	Yes	881 (0.08)	50 (0.14)	
Certain conditions originating in the perinatal period (P00-P96)	No	1094492 (99.68)	36612 (99.65)	0.467
	Yes	3540 (0.32)	127 (0.35)	
Congenital malformations, deformations and chromosomal abnormalities (Q00-Q99)	No	1091627 (99.42)	36481 (99.30)	0.004
	Yes	6405 (0.58)	258 (0.70)	
Injury, poisoning and certain other consequences of external causes (S00-T88)	No	1064455 (96.94)	35840 (97.55)	< 0.001
	Yes	33577 (3.06)	899 (2.45)	
Surgical Procedures				
Auditory surgery (CPT4: 69000-69979)	No	1096616 (99.87)	36698 (99.89)	0.401

	Yes	1416 (0.13)	41 (0.11)	
Cardiovascular surgery (CPT4: 33010-37799)	No	1092786 (99.52)	36413 (99.11)	< 0.001
	Yes	5246 (0.48)	326 (0.89)	
Digestive surgery (CPT4: 40490-49999)	No	1096760 (99.88)	36698 (99.89)	0.875
	Yes	1272 (0.12)	41 (0.11)	
Integumentary surgery (CPT4: 10030-19499)	No	1096378 (99.85)	36654 (99.77)	< 0.001
	Yes	1654 (0.15)	85 (0.23)	
Musculoskeletal surgery (CPT4: 20100-29999)	No	1096460 (99.86)	36705 (99.91)	0.014
	Yes	1572 (0.14)	34 (0.09)	
Urinary/Reproductive system surgery (CPT4: 50010-58999)	No	1095484 (99.77)	36540 (99.46)	< 0.001
	Yes	2548 (0.23)	199 (0.54)	

Table 3.3: USS Continuation

Variable	**Levels**	**No return visit N (%)**	**Had return visit N (%)**	**Chi-squared p value**
Demographics				
Age, y	[0, 5)	630446 (57.42)	24049 (65.46)	< 0.001
	[5, 12)	317500 (28.92)	8389 (22.83)	
	[12 or older)	150086 (13.67)	4301 (11.71)	
Sex	Female	510656 (46.51)	16632 (45.27)	< 0.001

	Male	587376 (53.49)	20107 (54.73)	
Payer	Commercial	256720 (23.38)	7967 (21.69)	< 0.001
	Medicare or Medicaid	631730 (57.53)	21966 (59.79)	
	Other governmental	24822 (2.26)	827 (2.25)	
	Self-Pay	66465 (6.05)	1987 (5.41)	
	Others	118295 (10.77)	3992 (10.87)	
Race and/or ethnicity	Caucasian	465203 (42.37)	15563 (42.36)	< 0.001
	African American	343185 (31.25)	11165 (30.39)	
	Hispanic	76179 (6.94)	2564 (6.98)	
	Asian Pacific Islander	13668 (1.24)	486 (1.32)	
	Native American	25517 (2.32)	1059 (2.88)	
	Other	174280 (15.87)	5902 (16.06)	
Respiratory conditions				
Acute nasopharyngitis [common cold] (J00)	No	1068164 (97.28)	35829 (97.52)	0.005
	Yes	29868 (2.72)	910 (2.48)	
Acute sinusitis (J01)	No	1089198 (99.20)	36531 (99.43)	< 0.001
	Yes	8834 (0.80)	208 (0.57)	
Acute pharyngitis (J02)	No	872131 (79.43)	30176 (82.14)	< 0.001

	Yes	225901 (20.57)	6563 (17.86)	
Acute tonsillitis (J03)	No	1048445 (95.48)	35385 (96.31)	< 0.001
	Yes	49587 (4.52)	1354 (3.69)	
Acute laryngitis and tracheitis (J04)	No	1096471 (99.86)	36686 (99.86)	0.972
	Yes	1561 (0.14)	53 (0.14)	
Acute obstructive laryngitis [croup] and epiglottitis (J05)	No	1038772 (94.60)	34513 (93.94)	< 0.001
	Yes	59260 (5.40)	2226 (6.06)	
jAcute upper respiratory infections of multiple and unspecified sites (J06)	No	683091 (62.21)	23327 (63.49)	< 0.001
	Yes	414941 (37.79)	13412 (36.51)	
Influenza due to certain identified influenza viruses (J09-J11)	No	1057642 (96.32)	35543 (96.74)	< 0.001
	Yes	40390 (3.68)	1196 (3.26)	
Viral pneumonia, not elsewhere classified (J12)	No	1096551 (99.87)	36649 (99.76)	< 0.001
	Yes	1481 (0.13)	90 (0.24)	
Bacteria and Other Pneumonia(J13-J18)	No	1044872 (95.16)	34173 (93.02)	< 0.001
	Yes	53160 (4.84)	2566 (6.98)	
Acute bronchitis (J20)	No	1073771 (97.79)	36112 (98.29)	< 0.001
	Yes	24261 (2.21)	627 (1.71)	
Acute bronchiolitis (J21)	No	1038129 (94.54)	33336 (90.74)	< 0.001
	Yes	59903 (5.46)	3403 (9.26)	
Vasomotor and allergic rhinitis (J30)	No	1073407 (97.76)	36226 (98.60)	< 0.001
	Yes	24625 (2.24)	513 (1.40)	

Chronic rhinitis, nasopharyngitis and pharyngitis (J31)	No	1094025 (99.64)	36647 (99.75)	< 0.001
	Yes	4007 (0.36)	92 (0.25)	
Chronic sinusitis (J32)	No	1081559 (98.50)	36290 (98.78)	< 0.001
	Yes	16473 (1.50)	449 (1.22)	
Chronic diseases of tonsils and adenoids (J35)	No	1095590 (99.78)	36680 (99.84)	0.015
	Yes	2442 (0.22)	59 (0.16)	
Nasal Polyp Other Unspecified Disorders Nose Nasal Sinuses(J33-J34)	No	998423 (90.93)	33687 (91.69)	< 0.001
	Yes	99609 (9.07)	3052 (8.31)	
Peritonsillar abscess (J36)	No	1097063 (99.91)	36664 (99.80)	< 0.001
	Yes	969 (0.09)	75 (0.20)	
Other diseases of upper respiratory tract (J39)	No	1096467 (99.86)	36698 (99.89)	0.1387
	Yes	1565 (0.14)	41 (0.11)	
COPD (J40-J44)	No	1073593 (97.77)	36053 (98.13)	< 0.001
	Yes	24439 (2.23)	686 (1.87)	
Asthma (J45)	No	896664 (81.66)	30209 (82.23)	0.006
	Yes	201368 (18.34)	6530 (17.77)	
Acute respiratory distress syndrome (J80)	No	1096515 (99.86)	36688 (99.86)	1.000
	Yes	1517 (0.14)	51 (0.14)	
Suppurative and nectrotic conditions of lower respiratory tract (J85-J86)	No	1097105 (99.92)	36710 (99.92)	0.791
	Yes	927 (0.08)	29 (0.08)	

Pleural effusion, plague, and other pleural conditions (J90-J92, J94)	No	1096757 (99.88)	36686 (99.86)	0.140
	Yes	1275 (0.12)	53 (0.14)	
Intraoperative and postprocedural complications and disorders not classified elsewhere (J95)	No	1096176 (99.83)	36638 (99.73)	< 0.001
	Yes	1856 (0.17)	101 (0.27)	
Other diseases of the respiratory system (J98-J99)	No	1082079 (98.55)	36175 (98.46)	0.203
	Yes	15953 (1.45)	564 (1.54)	

Table 3.4: Results of the Mixed Effects Models(MEM)

Variable	**Levels**	**Odds Ratio**	**p value**
ED Length of stay (hours)	[0,4)	Reference	
	[4,8)	1.131 (1.091, 1.173)	< 0.001
	[8,16)	1.149 (1.070, 1.235)	
	[16,24)	1.304 (1.181, 1.440)	
Previous ED Visit	0	Reference	
	1	1.367 (1.332, 1.403)	< 0.001
	2	1.606 (1.551, 1.663)	
	3 or more	2.244 (2.161, 2.331)	
Previous hospitalizations (prior 6 month)	Yes	1.202 (1.152, 1.255)	< 0.001
Has History Of Return Visit	Yes	1.158 (1.114, 1.204)	< 0.001
Number of medications greater than 90th percentile	Yes	1.399 (1.33, 1.471)	< 0.001
Season	Winter	Reference	
	Spring	0.988 (0.961, 1.015)	0.381
	Summer	1.014 (0.982, 1.047)	0.393
	Fall	0.928 (0.902, 0.955)	< 0.001
Comorbid Diagnoses			

Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism (D50-D89)	Yes	1.706 (1.528, 1.905)	< 0.001
Pregnancy, childbirth and the puerperium (O00-O9A)	Yes	1.511 (1.127, 2.025)	0.006
Neoplasms (C00-D49)	Yes	1.385 (1.100, 1.744)	0.006
Diseases of the circulatory system (I00-I99)	Yes	1.241 (1.080, 1.427)	0.002
Diseases of the genitourinary system (N00-N99)	Yes	1.233 (1.131, 1.343)	< 0.001
Mental, Behavioral and Neurodevelopmental disorders (F01-F99)	Yes	1.184 (1.083, 1.296)	< 0.001
Diseases of the digestive system (K00-K95)	Yes	1.134 (1.077, 1.194)	< 0.001
Diseases of the skin and subcutaneous tissue (L00-L99)	Yes	1.120 (1.050, 1.194)	0.001
Endocrine, nutritional and metabolic diseases (E00-E89)	Yes	1.112 (1.019, 1.214)	0.017
Certain infectious and parasitic diseases (A00-B99)	Yes	1.102 (1.062, 1.143)	< 0.001
Diseases of the nervous system (G00-G99)	Yes	1.100 (1.026, 1.179)	0.007

Injury, poisoning and certain other consequences of external causes (S00-T88)	Yes	0.810 (0.754, 0.870)	< 0.001
Diseases of the eye and adnexa (H00-H59)	Yes	0.804 (0.779, 0.830)	< 0.001
Surgical Procedures			
Integumentary surgery (CPT4: 10030-19499)		1.832 (1.461, 2.296)	< 0.001
Urinary/Reproductive system surgery (CPT4: 50010-58999)		1.603 (1.381, 1.860)	< 0.001
Cardiovascular surgery (CPT4: 33010-37799)		1.491 (1.324, 1.678)	< 0.001
Digestive surgery (CPT4: 40490-49999)		0.725 (0.526, 0.999)	< 0.001

Table 3.5: MEM Continuation

Variable	**Level**	**Odds ratio (95% CI)**	**p value**
Age, y	[0, 5)	Reference	
	[5, 12)	0.782 (0.760, 0.804)	< 0.001
	[12 or older)	0.816 (0.786, 0.848)	
Payer	Commercial	Reference	
	Medicare or Medicaid	1.127 (1.092, 1.163)	< 0.001
	Other governmental	1.142 (1.059, 1.232)	< 0.001

	Self-Pay	1.042 (0.988, 1.098)	0.128
	Others	1.202 (1.149, 1.257)	< 0.001
Race and/or ethnicity	Caucasian		
	African American	0.940 (0.912, 0.969)	< 0.001
	Hispanic	0.964 (0.919, 1.011)	0.134
	Asian Pacific Islander	1.054 (0.959, 1.159)	0.274
	Native American	0.928 (0.835, 1.031)	0.165
	Other	1.017 (0.982, 1.054)	0.339
Sex	Female	Reference	
	Male	1.025 (1.003, 1.047)	0.024
Respiratory risk factors			
Peritonsillar abscess (J36)	Yes	2.266 (1.775, 2.892)	< 0.001
Viral pneumonia, not elsewhere classified (J12)	Yes	1.436 (1.155, 1.785)	0.001
Acute bronchiolitis (J21)	Yes	1.390 (1.332, 1.449)	< 0.001
Bacteria and Other Pneumonia(J13-J18)	Yes	1.273 (1.216, 1.333)	< 0.001
Other respiratory conditions			
Acute pharyngitis (J02)	Yes	0.948 (0.915, 0.982)	0.003
Asthma (J45)	Yes	0.924 (0.894, 0.956)	< 0.001

Acute upper respiratory infections of multiple and unspecified sites (J06)	Yes	0.910 (0.885, 0.936)	< 0.001
Acute nasopharyngitis [common cold] (J00)	Yes	0.907 (0.845, 0.975)	< 0.001
Acute tonsillitis (J03)	Yes	0.896 (0.844, 0.950)	< 0.001
COPD (J40-J44)	Yes	0.888 (0.820, 0.961)	0.003
Influenza due to certain identified influenza viruses (J09-J11)	Yes	0.881 (0.828, 0.939)	< 0.001
Nasal Polyp Other Unspecified Disorders Nose Nasal Sinuses(J33-J34)	Yes	0.875 (0.840, 0.911)	< 0.001
Chronic sinusitis (J32)	Yes	0.859 (0.779, 0.946)	< 0.001
Acute bronchitis (J20)	Yes	0.819 (0.754, 0.890)	< 0.001
Acute sinusitis (J01)	Yes	0.768 (0.667, 0.884)	< 0.001
Chronic diseases of tonsils and adenoids (J35)	Yes	0.724 (0.558, 0.939)	0.015
Vasomotor and allergic rhinitis (J30)	Yes	0.683 (0.624, 0.748)	< 0.001
Chronic rhinitis, nasopharyngitis and pharyngitis (J31)	Yes	0.665 (0.540, 0.820)	< 0.001
Intraoperative and postprocedural complications and disorders not classified elsewhere (J95)	Yes	0.641 (0.491, 0.837)	0.001

Table 3.6: Performance of Machine Learning Models

Model Name	AUC Values
DCT	0.514
KNN	0.532
GNB	0.535
MNB	0.532
CNB	0.546
MLR	0.548
RRC	0.546
LCSGD	0.553
PAC	0.534
SVC	0.544
RF	0.546
ERT	0.542
GBT	0.558
EGBT	0.645
MLP	0.519

4 General Pediatric Models for Understanding and Predicting Prolonged Hospital Length of Stay

The impact of the deterioration in the health of a child requiring hospitalization on the family is multifaceted. It encompasses social, economic, and psychological impressions. To alleviate those burdens, considerable efforts have been deployed, with the ultimate goal of improving the healthcare system, the organizational management of modern hospitals and improving the quality of care delivered to patients. Some pediatric hospitals address these concerns by creating departments such as Child Life to improve pediatric patients, especially that of the young child. The length of stay (LOS) in the emergency department (ED) patients has been found to be a top priority for hospitals and health systems.

Traditionally, ED length of stay (LOS) has been used as an accurate metric to assess the efficiency of ED management, patient quality, patient quality of care, and functional evaluation [78], [79]. In fact, it has generally been accepted that shorter LOS is associated with more efficient and effective care [80]. Based on the existing standards, it has been suggested that an appropriate median ED holding time should be of 2.0 hours for a hospital bed (1.5 hours for an intensive care unit (ICU)) [81]. For [82], patients presenting to the ED should receive a decision in a maximum of 6 hours after admission to the ED and leave ED at this time. The shorter a patient stays in the ED, the smaller the chance of developing infections is [79]. Also, reduced ED LOS is associated with decreased mortality, drastic reduction of the social costs, and financial burdens not only on patients but also on medical services [83], [84]. Furthermore, reduced LOS can avoid unnecessary expenses and free up beds for other critically ill patients.

In contrast, longer ED LOS can result in exposing patients to serious healthcare-acquired infections, higher mortality rates, and imposing noticeable increase in the total social and

financial related costs. The impact of prolonged length of stay transcends the child's experience to that of the family and the hospital facility itself. Longer length of stay may result in reduced working hours of a parent or guardian, lost productivity at work, and increased psychological and financial burden on the family. Parents struggle to pay for prolonged care, which has an impact on their current and future financial security that in turn affects their children. A 3-year study of hospital resource utilization related to childhood cancer from 2005 found the cumulative charges reaching $16 million, which is about $100,000 per child in the study [85]. In addition, 50% of those charges were in the first four and a half months of diagnosis and those charges surprisingly were distributed among only 12.7% of patients [85]. These patients had worse diagnoses and underwent a multitude of treatments. Parents of children with a cancer diagnosis, even with insurance, accrue a significant bill to pay for the treatment. The children themselves are likely to endure difficult mental health and physical challenges as their LOS increases. In addition, providers and hospital facilities are more likely to experience revenue leakage as well as challenges with adjudicating medical claims for treatment offered. It is therefore imperative for the hospitalized child, the family, providers, and the hospital to reduce unnecessarily prolonged LOS.

One of the most significant issues of ED prolonged LOS is overcrowding. ED overcrowding has recently become a worldwide concern. ED overcrowding may result in creating many problems patients and staff, increasing the waiting times and the ED resource utilization, augmenting the length of stay, portending to an increase of errors and patient mortality, and inflicting serious financial losses to hospitals [86], [87], [88].

Several studies have been conducted to determine risk factors of longer LOS. Internal and external factors such as patient characteristics, ED staffing, health care providers, time of patients arrival, diagnostic methods as well as hospital resource utilization, allocation, and administration have strongly been associated with ED prolonged length of stay [81], [89], [87], [79]. The impact of malnutrition [90], hyponatremia [91], febrile neutropenia [92], clinical

pathways [84], [93] and weekend admission [94] on hospital length of stay are among the studies conducted with the ambition to well understand eventual factors that contribute to ED process times and patient care delays. An 18-year study from 2018 on a healthcare center in Mexico assessed several sociodemographic and disease-specific differences for a patient's LOS [79] Oncology patients often require more hospitalizations during treatment than patients with most other conditions. A study on the LOS of cancer patients in Brazil explored the association between demographics and clinical attributes with LOS within the first year of outpatient treatment [95].

In this study, we addressed the identification of novel risk factors for prolonged LOS using advanced statistical analysis and the prediction of patients most likely to experience prolonged LOS using machine learning on a large multicenter electronic medical records database. The goal of this study is to use data captured during the first 24 hours of admission to predict patients with LOS greater than 2 weeks. The first 24 hours of hospitalization of a child has yielded critical information to predict LOS [75]. Obviously, appropriate LOS depends on diagnoses, severity of illness, and a host of complex clinical considerations. We chose a 2-week threshold because it is much greater than the average pediatric length of stay and falls between the 90th and 95th percentile of general pediatric LOS.

4.1 Methods

Dataset, Clinical Admission, and Patient Encounter

This investigation was approved by the Institutional Review Board of CHOC Children's (#180857). A retrospective, population-based cohort study using the Cerner Health Facts Database for the US-based patients is conducted. The Cener Health Facts database is a de-identified patient database that provides records about patients coming from different participating health institutions. The database contains time-stamped and sequenced infor-

mation on pharmacy, laboratory, admission and billing data from all patient care locations. In addition, Cener Health Facts database captures information demographic, hospital admissions and discharge, and diagnostic procedures. As of 2018, the Health Facts database was composed of records of more than 65 million patients provided by over 100 US healthcare systems and over 650 facilities, more than 500 million encounters, approximately more than 500 million encounters, and 4.7 billion laboratory results. For the sake of confidentiality and compliance with HIPAA privacy regulations, patient names are omitted and a unique identification number is assigned to each single patient, which is systematically available to all the participating health institutions. Structurally, Cener Health Facts database stores and retrieves data that is represented in smaller databases or tables. The database is available to researchers at healthcare systems that contribute data to it. In 2020, the database was upgraded and reconstituted as the Cerner Real World Data [96].

We retrieved all pediatric encounters (patients less than 18 years) from the database and excluded encounters that occurred during the first 6 months of the very first encounter for each hospital. This exclusion was necessary to ensure that encounters in each hospital have a 6-month history for variables capturing patients' history. Furthermore, we excluded all encounters with LOS less than 24 hours to ensure that all qualifying encounters had a minimum of 24 hours LOS and data within the first 24 hours of the admission. An additional exclusion by hospital facility was carried out by excluding all facilities with less than 1000 encounters. The qualifying hospitals contributed data from different periods between 2001 and 2017. We calculated the average age of patients seen at each hospital and classified those less than 18 years as free-standing pediatric hospitals.

We retrieved data on patient demographics, payer, medications administered during the first 24 hours of admission, type of admission, and previous healthcare utilization 6 months prior to the encounter start date. The variable race/ethnicity was divided into 6 groups: Caucasian, Hispanic, Black/African American, Asian/Pacific Islander, Native American and

Other/Unknwon. The type of insurance covering the patients were: commercial, governmental (Medicare, Medicaid), other overnmental (Champus, etc), self pay, and otner. All International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes in the database were converted to the Tenth Revision (ICD-10-CM) using appropriate translation tables and merged with the original ICD-10-CM data from the most recent encounters. In order to automatically capture the largest number of the most prevalent pediatric conditions as variables for the study, we retrieved the three-digit level of all ICD-10 diagnoses. This, however, introduced very rare encounters that may result in issues with statistical separability as well as impact the value of corresponding findings. Consequently, we set an a priori threshold for including a three-digit level diagnosis at an incidence rate of 1%. In other words, only such diagnoses occur in 1% or greater was included. These considerations also help in reducing the risk of modeling challenges due to statistical class separability which may result in inflated effect sizes or inestimable parameters [97]. We defined the outcome variable, prolonged LOS, as any LOS greater than 2 weeks.

Outliers in continuous variables may negatively affect statistical and machine learning models. We handled outliers excluding by records wherein the value of a continuous variable was greater than the 99.5th percentile. We checked for multicollinearity between variables by estimating the generalized variance inflation factor (GVIF) that measures the degree to which correlation between variables in a model result in inflated variances [98] We set an a priori threshold of 4 for the GVIF wherein all variables with GVIF greater than the threshold is removed in a stepwise manner until all variables have GVIF less than 4 [75] We split the data into two equal halves by patients.

Statistical Analysis

A two-stage statistical analysis study is implemented using ED LOS as outcome variable of interest. First, we descriptively described the data via a summary statistics. Second, we built

a statistical model for inference using a mixed effects logistic regression model where hospitals were modeled as random intercepts with patients nested within them. Mixed models are an extension of the traditional linear models which at its core incorporate random and fixed effects. Mixed models (MM) are well suited describing analysis of clustered or longitudinal nominal or ordinal response data [99]. They are most useful when working with hierarchical data which often exude some degree of dependency (within-subjects designs /longitudinal data). An example would be subjects being sampled from within doctors. MM systematically accounts for within subjects variability thus, working around ANOVA assumption that data points are independent of one another [100]. MM are also efficient at analyzing data that are non-independent and correlated. MM address the correlation issue and does not violate the linear model's assumption- independence of observations.

To put MM at work, we started with a full model with all variables as well as 2 two-way statistical interactions: (1) between age and diagnoses, and (2) between the number of medications and diagnoses. The selection of these interaction terms was inspired the need to capture difference in disease progression by age and by severity of illness. We selected the number of medications administered during the first 24 hours as a proxy for severity of illness since there are no standard measures of severity.

Given the fact numerous statistical models, such as mixed effects model, have serious problems with handling very complex and non-linear interactions among predictors themselves as well as with the target variable, it is worth the efforts to trying other robust models. Thus, to further the study, we used the extreme gradient boosting (XGBoost) to assess higher level interactions beyond the 2-way statistical interactions(for example 3 or more) considered in the mixed effects model. Extreme gradient boosting is a variant of of the Gradient Boosting Machine. Virtually, XGBoost aims at using an ensemble of weak learners that are sequentially trained in the sake of having a stronger classifier or regresser model. A meta-optimization task was performed by tuning some hyperparameters, which make the

ensemble training controllable. This was indeed done with cross validation technique for assessing how the results of a model will perform on an unseen data. Table 4.1 displays the values assumed by the involved hyperparameters.

Table 4.1: Parameter grid for hyperparameter tuning of the extreme gradient boosting model

Hyperparameter	Values	Significance
Boosting operations/iterations	64	Number of boosting operations is equivalent to the number of trees built. We chose 64 based on previous experience on the minimum number of boosting iterations required
Learning rate	0.2, 0.3, 0.5, 0.8	Relating to how fast the model learns. Smaller values help to prevent overfitting
Maximum tree depth	4, 6, 8	The depth of each tree which controls the complexity of the model and interactions explored
Minimum child weight	0, 1, 2, 4	Relating to how partitions are made on a child node. Larger values create more conservative models
Gamma	0, 1, 2, 4	Relating to how leaf node partitions with respect to changes in loss. Larger values result in more conservative models

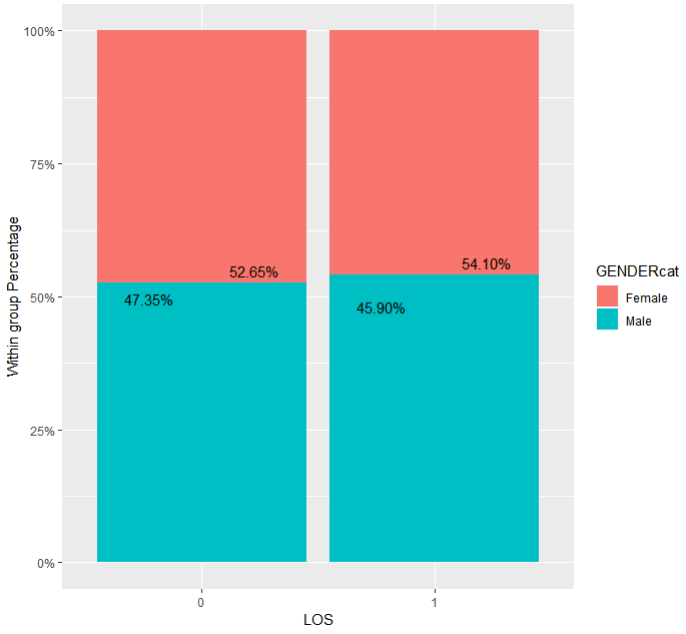


Figure 4.1: LOS with respect to patient's gender

Numerous metrics such as the area under the receiver characteristics curve (AUROC), the area under the precision-recall curve (AUCPR), and the values of model sensitivity, positive predictive value, negative predictive value, F1 score, and the number needed to evaluate (as the number of patients that will be flagged at-risk before a true positive prediction) were used to evaluate the performance of both the statistical mixed effects regression and machine learning models.

4.2 Results

Exploratory Data Analysis : Free-standing Pediatric and Mixed Medical Facilities

Sixty medical centers met the inclusion criteria for hospitals. Of the medical centers used in this study, nine (9) were free-standing pediatric hospitals, with the remainder being mixed

medical centers. The 9 free-standing pediatric hospitals contributed ∼51% of all encounters in the study thereby accounting for the longest period and the greatest number of daily pediatric admissions.

Exploratory Data Analysis: Length of Stay Class

A total of 700,00 patients with ∼1,000,000 encounters met the inclusion criteria for encounters. The rate of prolonged LOS was 5.0% across all patients and their encounters. The mean age of the patients was 5 years. Interestingly from the data, we observed that female patients account for ∼47.3% of LOS class 1 while male subjects were ∼54.1% of LOS class 2 (Figure 4.1). From the within group percentage of both LOS classes high cases of emergent admissions were observed, with LOS 1 and LOS 2 accounting for 83.7% and 85% of the total number of patients Figure 4.2.

Expectedly, the combination of commercial and governmental health insurance accounted for over 60% of all patients coverage plans while 2.2% and 1.8% of the patients paid out of pocket in both LOS classes (Figure 4.3). From the withing group percentage of patients, we observed that there were nearly as many children as adults in LOS 1 with children accounting for ∼50.4% of all patients, whereas there was a 3/2 ratio of children to adults in LOS 2, where adults account for ∼40.8% (Figure 4.4). Furthermore, we explored the ethnicity/race distribution of the patients. Notably, patients with Caucasian background accounted for approximately half of all races represented in both LOS classes. Patients of African descent accounted for ∼20% of all patients. Strikingly, ∼20% of the subjects' ethnicity were undisclosed while patients of Hispanic, Asian, Native American origin accounting for ∼4%, ∼2%, and ∼2% of all LOS 1 and LOS 2 subjects (Figure 4.5).

Descriptively (Table 4.5), we surprisingly found that apart from variables (11) including Overweight and obesity (E66), Scoliosis (M41), Pervasive developmental disorders (F84), Pneumonia, unspecified organism (J18), Other and unspecified soft tissue disorders, not

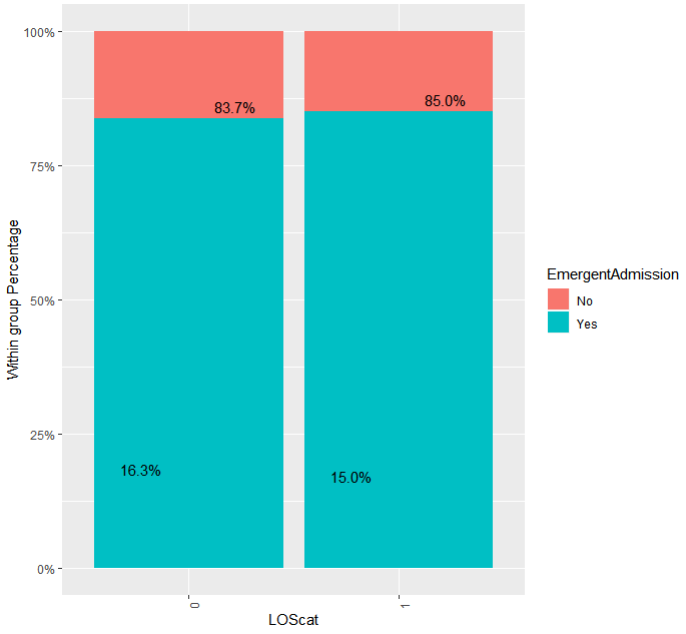


Figure 4.2: LOS with respect to Emergent Admission

elsewhere classified (M79), Other symptoms and signs involving the circulatory and respiratory system (R09), Fever of other and unknown origin (R50), Headache (R51), Liveborn infants according to place of birth and type of delivery (Z38), Family history of other specific disorders (Z83), and Personal risk factors, not elsewhere classified (Z91), there is existence a significant ($P < 0.001$) association between the prevalence of prolonged LOS in emergency department and the remaining variables which constitute the data set.

Multivariate Analysis : mixed effects model

In this retrospective multi-centered study, a mixed effects model (with hospitals as random intercepts and patients nested within hospitals) was constructed to evaluate the association between LOS and the involved independent variables in the ED settings. The results obtained from MM are shown in Table 4.6. It was found that a statistically significant association between such as variables patient demographics, healthcare utilization variables, certain diagnoses without interactions, certain interactions between age and diagnoses, and certain

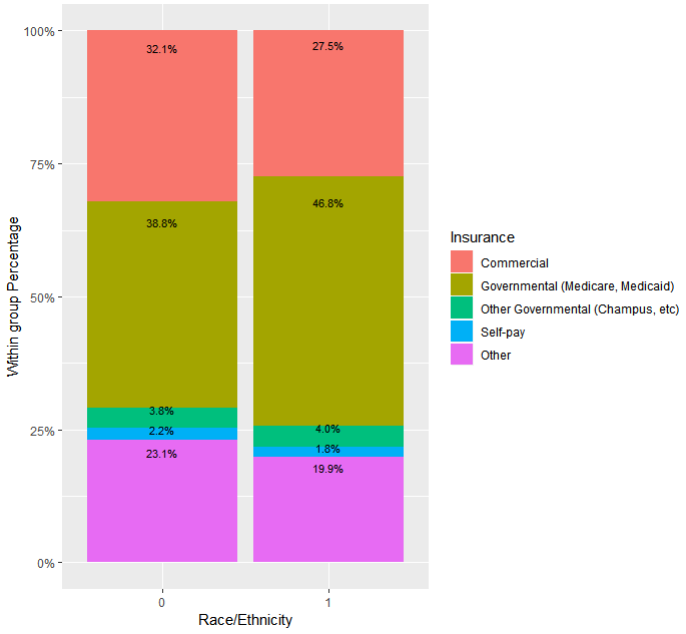


Figure 4.3: LOS with respect to insurance cover

interactions between number of medications and diagnoses, and prolonged length of stay exists.

Variables with no interaction(s)

Male patients were significantly more likely to have prolonged LOS with a 6.7% increase in the odds. Compared to the baseline group of patients that have commercial health insurance, patients that have governmental insurance (Medicare, Medcaid) or any other types of government insurances (Champus, etc) have higher odds of experiencing prolonged LOS with increases 35% and 21% respectively. Similarly, self-insured patients have 35% higher odds of prolonged LOS compared to commercially insured patients. Patients that possess other health insurance or coverage have 34% higher odds than the baseline group of patients. Compared to Caucasian, patients from other races/ethic groups have higher odds of experiencing prolonged LOS with increases of approximately 0.04-, 11-, 11.20-, 15% for Asian/Pacific Islander, Hispanic, Other/Unknown, and Native American respectively. We

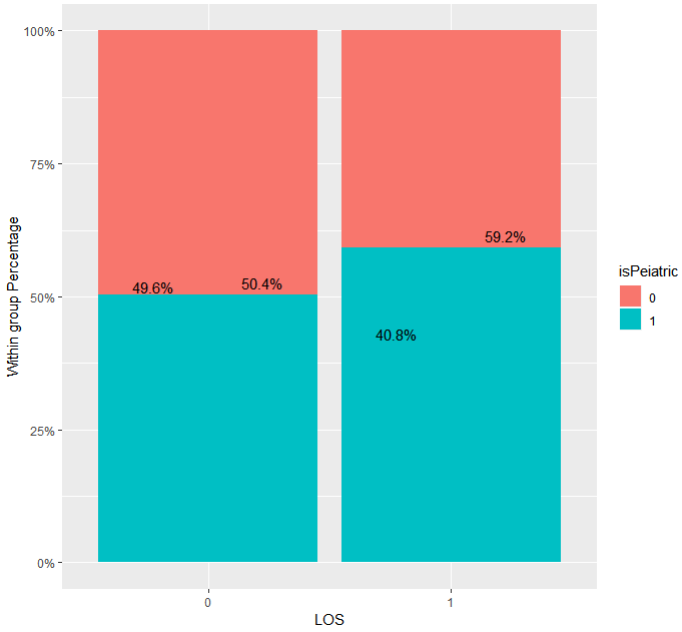


Figure 4.4: LOS with respect to patient's age

also found that patient's maximum of length of stay in the last 6 months, emergency admission status and past readmission history are significantly associated with the odds of extended length of stay in ED. Along with being with such significant associations, they all have an increase in odds of having a prolonged LOS ranging from approximately from 4% to and 67% with maximum previous length of stay(last 6 months) and emergency admission variables assuming respectively the smallest and largest increases in odds. In contrast, patient's history of ED visits has a decrease in odds of causing a prolonged length of stay. The Free-standing pediatric hospital variable appears to not be a significant factor to prolonged LOS.

Certain diagnoses were significantly associated with prolonged LOS. These diagnoses include Lymphoid leukemia (C91), Type 1 diabetes mellitus (E10), Suppurative and unspecified otittis media(H66), Pneumonia, unspecified organism (J18), Disorders of newborn related to short gestation and low birth weight, not elsewhere classified (P07), Family history of certain disabilities and chronic diseases (leading to disablement) (Z82), Personal history of

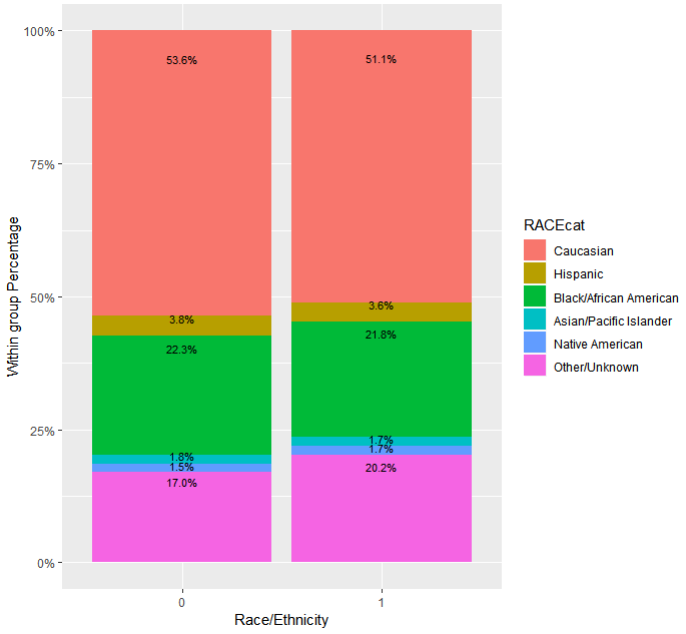


Figure 4.5: LOS with respect to race/ethnicity

other diseases and conditions (Z87), Pervasive developmental disorders (F84), and Sleep disorders (G47). The decrease in odds associated with these conditions are greater than 18- and less than 72%. On the other hand, the remaining conditions are risk factors that are highly associated with extended length of stay with an increase of odds ranging from 3% to 445%. **Variables with interaction(s)**

An interaction arises if there is an eventual relationship among two or more variables. [101] defines a statistical interaction as a scenario where the relation between a predictor and a target variable depends on another independent variable, named as a moderator. Our mixed effects model gives us a mathematical model that we can we use to estimate the probability of a patient stay more than expected in the hospital given certain independent variables. For example, suppose that we have three predictors, X, Z, and XZ. The model could roughly be expressed as:

$$log(\frac{\pi}{1-\pi}) = \alpha_0 + \alpha_1 X + \alpha_3 Z + \alpha_3 XZ \tag{4.1}$$

wherein α_0, α_1, α_2, α_3 are the coefficients of the model. The statistical significance of the interaction coefficient α_3 informs about the association between X and the probability that the outcome variable Y be equal to 1 depends on the predictor Z.

Table 4.8 and Table 4.9 show the results of two separate multivariate analyses. First, we chose to discuss the results from the interaction effects between two main effects and the remaining independent variables in the model: one for age of patient and the other one is for the number of medications administered to a patient. In this process, we not only examine the significance of the risk factor(s), but also the strength of the association. Then, the odd ratios is calculated for each effect present in the model. As mentioned earlier, this is conducted by deploying a mixed effects model.

Statistical interactions with age

Table 4.8 describes the main effects of age and those of the other existing independent variables. The results show that each main effect in the model is statistically and significantly ($P < 0.001$) associated with prolonged LOS. Table 4.8 also shows that the interaction between the factor age and any other factor is strongly associated with staying in ED more than 2 weeks. In addition, we notice that the main effects Age, Pervasive developmental disorders (F84), Asthma (J45), Other disorders of urinary system (N39), Sleep disorders (G47), Convulsions, Not elsewhere classified (R56), and Obstructive and reflux uropathy (N13), with negative coefficients are less likely to be important risk factors of prolonged length of stay. This tells us that in the absence of statistical interactions with medical conditions a patient is undergoing, older patients are less likely to have prolonged length of stay. In contrast, without any interaction, patients that suffer from Streptococcus, Staphylococcus, and Enterococcus as the cause of diseases classified elsewhere (B95), Other anemias (D64), Major

depressive disorder, single episode (F32), and Attention-deficit hyperactivity disorders (F90) are more likely to stay longer in hospital in the case their condition worsens.

From the second part of Table 4.8, we can easily see that the interactions between age and conditions such as Pervasive development disorders (F84), Asthma (J45), Certain disorders of the urinary system(N39), Streptococcus, Staphylococcus, and Enterococcus diseases classified elsewhere (B95), Certain functional intestinal disorders (K59), and Sleep disorders (G47) are statistically significant to the prolonged length of stay in hospital. In addition, we can notice that these conditions interacting with age produce some positive effects on the length of stay. This implies that patients that face the deterioration of their health, based on the evolution of these conditions, are more likely to stay longer than expected in hospital.

For the sake of understanding the concept of 2-way statistical interaction between age and diagnoses or medical conditions, let us consider some graphical representations. Do note that age and diagnoses are a continuous variable and a dichotomous variable respectively. Thus, the two-way interaction between age and a particular diagnosis fit separate regression lines (on the log-odds scale) for the effects of age for the groups with and without the condition. A positive/negative interaction effect reflects a higher/lower slope for the line depicting the effect of age for the group with the condition compared to the group without. The intercepts of the lines are functions of the main effects that in combination with the slopes can produce different scenarios (non-parallel non-intersecting lines within the range of values of age with various slopes or non-parallel intersecting lines within the range of values of age with various slopes). These scenarios are displayed (on the probability scale).

Figure 4.6 displays the statistical 2-way interaction between age and 4 diagnoses with positive effects. The plots in Figure 4.6 show the predicted probability of prolonged length of stay with respect to patients' age. In the top left plot, we observe that the directions of the effects size of the age between the level of the diagnosis are opposite. Second, we notice the

presence of a statistical interaction between age and the Pervasive development disorders (F84). Such an interaction is sustained by the intersection point between the two graphs. Third, we can deduce that younger patients that do not suffer from Pervasive development disorders (F84) are more likely to have a prolonged length of stay than patients that do not have such a condition. Fourth, we remark that older patients with this condition have longer length of stay. The right top plot is about two non-parallel lines depicting the level of the asthma diagnosis. Since the two lines are not parallel, we can say that they will certainly intersect at a certain age greater than 17. As with the previous graph, we can see that the direction of the effects size of the age among patients with asthma and those with no-asthma are opposite. It results from this that older patients with asthma conditions have higher chance to have a prolonged length of stay. The left bottom plot shows that patients with certain disorders of the urinary system(N39) are more likely to have a longer length of stay once they get older. The right bottom graph indicates that the interaction associated with infections due to Streptococcus, staphylococcus, and enterococcus(B95) was insufficient to change the trend imposed by the main effects of age even though the direction of the effect size of the interaction term was positive.

Statistical interactions could also have negative effect sizes. The second part of Table 4.8 shows some examples where the coefficients of condition-age interaction are negative values. Arranged in decreasing order of the magnitude, the involved conditions in our study that have negative interaction effect sizes with age include Attention-deficit hyperactivity disorders (F90), Major depressive disorder, single episode (F32), Obstructive and reflux uropathy (N13), Convulsions, not elsewhere classified (R56), and Other anemias (D64). Figure 4.7 represents the plots of some statistical interactions(with negative interaction effect sizes) between age and the top 4 conditions. In the top left graph, the predicted probability length of stay is displayed in terms of the age of the patients. From this graph, we can see that the direction of the effects size of the age among patients that suffer from attention-deficit

hyperactivity disorders (F90) condition and those who do not have the so-called condition are opposite. We can then notice an existing interaction between the two level of the attention-deficit hyperactivity disorders (F90) condition. We can also notice that younger patients with the attention-deficit hyperactivity disorders (F90) condition have higher probability of having longer length of stay. A reverse scenario is seen with older patients. Similar findings are displayed in the top right graph, with the exception that among the patients that do not have the conditions, the predicted probability of longer LOS is much smaller in this case. The bottom two graphs present the almost the same features in the direction of direction of the effects size of the age among patients that suffer from Obstructive and reflux uropathy (N13) and Convulsions, not elsewhere classified (R56). However, it should be highlighted that patients with these two mental health conditions tend to have longer length of stays than their peers. The graphs were interpolated cutting across ages younger than the minimum age for diagnosis of these conditions. Obstructive and reflux uropathy (N13) and convulsions (R56) had similar interactions with patients diagnosed with these conditions have reduced odds for prolonged length of stay and further reduction in odds as patient age increases.

Statistical interactions with the number of medication administered during the first 24 hours of hospitalization

Table 4.9 depicts the statistical interactions between the number of medications and the other independent variables that are present in our model. It results from the first part of the table that besides the factors Gastro-esophageal reflux disease (K21), Other functional intestinal disorders (K59), Scoliosis (M41), Neonatal jaundice from other and unspecified causes (P59), Transitory disorders of carbohydrate metabolism specific to newborn (P70), lack of expected normal physiological development in childhood and adults (R62) and Personal risk factors, not elsewhere classified (Z91), which accounts for of all the main effects in the model, the remaining factors are significantly and statistically($P < 0.001$) associated with prolonged length of stay.

The second part of Table 4.9 indicates the interaction effects between the number of medications and diagnoses. Strikingly, we found statistically significant interactions between the number of medications and the factor diagnoses present in the study. Of these statistically significant interactions, ~ 20 percent have a negative interaction effect on the length of stay.

To better illustrate the effects of the 2-way statistical interactions between the number of medications administered and different diagnoses, let us analyze the corresponding graphical representations. The scenarios are identical for the interaction between the number of medicines that have been administered to the patients during the first 24 hours after admission and particular diagnoses. On the log-odds scale, the two-way interaction between the number of medications and a diagnosis fits separate regression lines for the effects of number of medications for the groups with and without the condition. A positive/negative interaction effect reflects a higher/lower slope for the line depicting the effect of age for patients under medications compared to the ones without medications. As previously stated, the intercepts of the lines are functions of the main effects that in combination with the slopes can produce different scenarios (non-parallel non-intersecting lines within the range of values of age with various slopes or non-parallel intersecting lines within the range of values of age with various slopes). Figure 4.8, and Figure 4.9 depict these scenarios on the probability scale.

For this, we consider 4 top (diagnoses Liveborn infants according to place of birth and type of delivery (Z38), Disorders of newborn related to long gestation and high birth weight (P08), Viral and other specified intestinal infections (A08), and Acute bronchiolitis (J21)) that have positive interaction diagnoses with number of medication factor (Figure 4.8). In the top left graph where the predicted probability length of stay is displayed in terms of the number of administered medications. We notice that with respect to small number of medications administered in the first 24 hours after admission, newborns and children of other ages (who were not born during the index encounter) have almost the same odds of prolonged length

of stay. Opposed to this, as the number of administered medications increases within 24 hours, the odds for newborns exponentially increases. We also see that after 25 administred medications, the odds for longer length of stay remains the same among liveborns. This is indicated by the existence of the plateau reached after 25 administered medications. A similar but less pronounced effect is seen for disorders of newborns related to the remaining top 3 diagnoses.

Lastly, in the context of negative interaction effects, we chose the top 4 diagnoses which are congenital malformations of cardiac septa (Q21), scoliosis (M41), congenital malformations of the great arteries (Q25), and essential primary hypertension (I10). Figure 4.9 gives the graphical representations of the corresponding interactions with the number of administered medical. First, we notice that for smaller number of medications administered within 24 hours, regardless of the age, the patients have the same likelihood for longer length of stay. In addition, it is shown that the effects size of the number of administered with 24 hours among liverborns and among patients of other ages is the same are in the same direction. It is also found patients with these conditions have reduced odds of prolonged length of stay but the odds increases as the number of medication increase

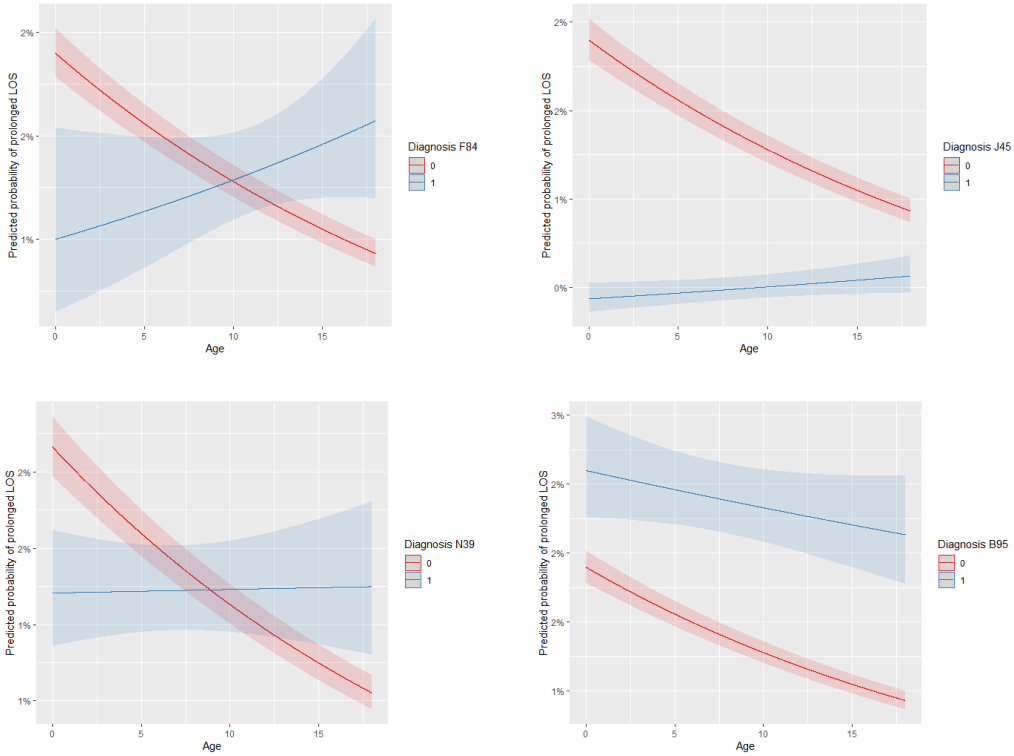


Figure 4.6: Positive interaction effect sizes with age

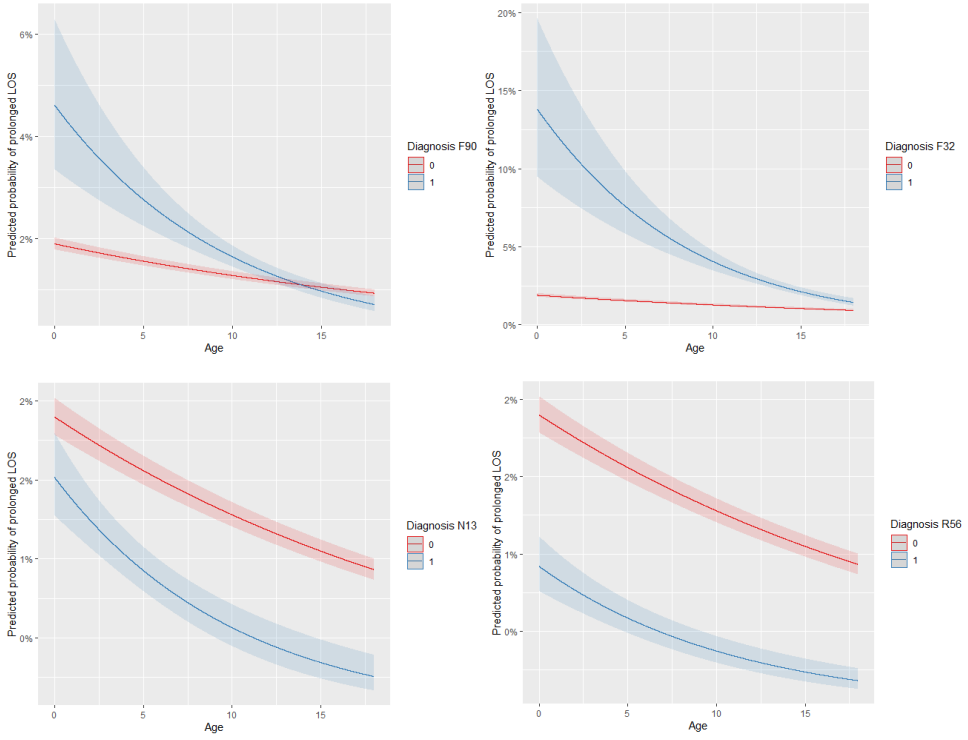


Figure 4.7: Negative interaction effect sizes with agewith age

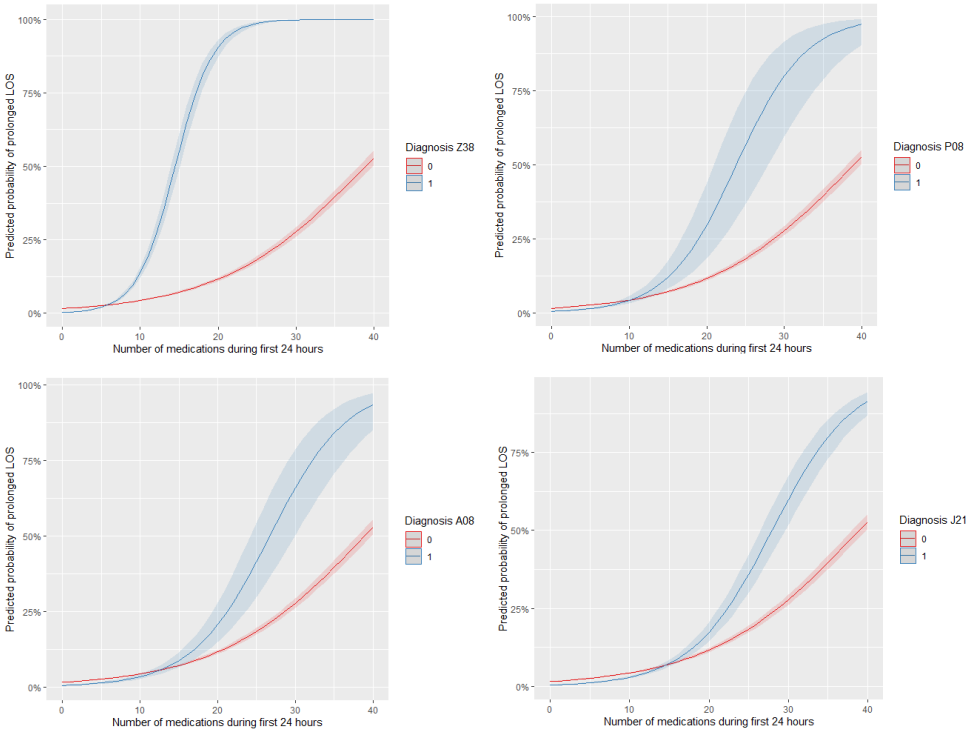


Figure 4.8: Positive interaction effect sizes with number of medications

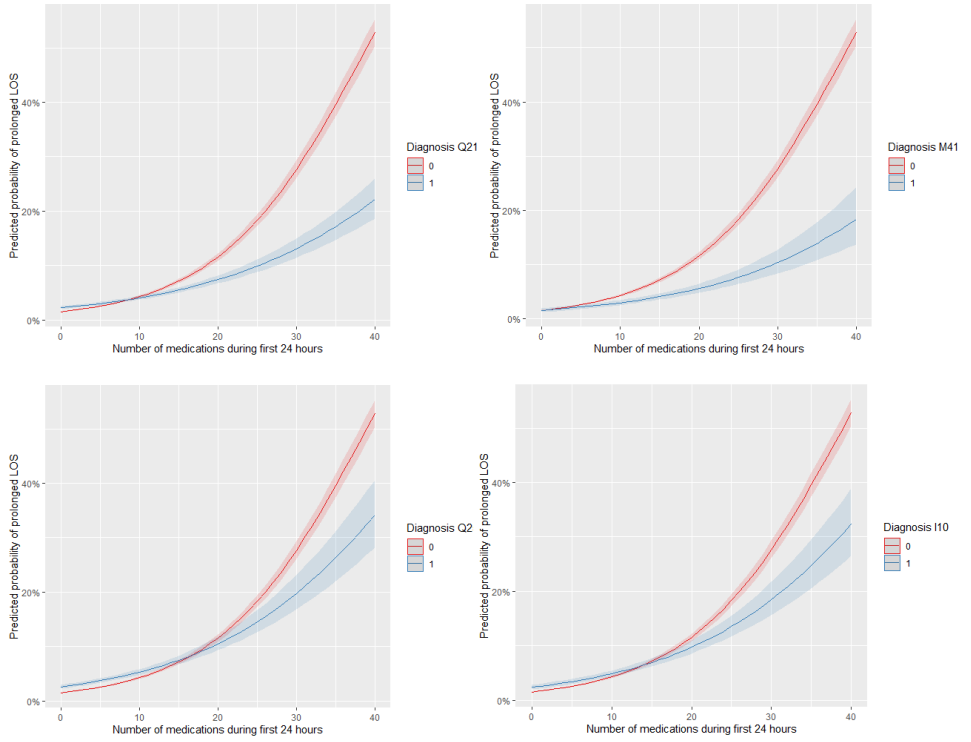


Figure 4.9: Negative interaction effect sizes with number of medications

Machine Learning Consideration

The amenable meta-optimization task achieves certain parameters that can be assumed as the best performing set of parameters. In the best case, we have a maximum tree depth of 8, learning rate of 0.3, gamma of 4, and minimum child weight of 0. A full description of the involved parameters is given in Table 4.1.

Model Performance

As mixed effects model returns a probability, a classification threshold of value 0.068 was set since the beginning of the study. A value greater or equal to 0.068 will indicates that a specific patient is at high risk of having a prolonged length of stay, a value below indicates that the patient is not at high risk of staying more than expected. At this threshold and in the absence of interventions, of 4 patients that are flagged as being at high risks only one of them would have a prolonged length of stay. With the set threshold, the mixed effects model achieved an AUROC of 0.887 (0.885, 0.889), and at a specificity of 0.900, the sensitivity positive, positive predictive value, F1 score, and NNE were 0.667, 0.264, 0.380 and 3.782 respectively. The area under the precision recall curve was 0.513.

On the other hand, we observed a much higher performance of the chosen machine learning model. The AUROC obtained by the extreme gradient boost model was 0.931 (0.930, 0.933). At a specificity 0.900, the other metrics including sensitivity, positive predictive value, F1 score, and NNE assumed 0.786, 0.294, 0.428, and 3.398. An interpretation of this obtained NNE values indicates that at the set threshold, 2 in 7 patients flagged at high risk for prolonged length of stay would indeed have a prolonged length of stay in the absence of interventions. The are under the precision-recall curve was 0.611.

4.3 Discussion

In this work, a retrospective study was conducted to identify all of the clinically vital factors that are statistically and significantly associated with hospital length of stay. For this, a mixed effects model was used to extensively analyze the factors as well as a machine learning algorithm for prediction. In this stride, interesting findings were observed. Some of the findings are in full agreement with what clinically cited as factors to hospital prolonged length of stay, and others are newly discovered. Newborns and neonatal population tend to have the longest length of stay. These patients typically stay longer than 2 weeks. This is expected in the case of admissions to the NICU after birth and within the first 28 days of life. This finding reinforces the idea that neonatal population has very distinctive needs and clinical presentation than the rest of the pediatric population. The proportion of neonatal children with prolonged length of stay is 0.087 and significantly greater than their older peers with a rate of 0.040. This creates a strong case for follow-up studies with focus specifically on the neonatal population to find risk factors peculiar to neonates of which interventions for improved quality of care may be developed.

The next set of findings is related to pediatric infectious diseases, cerebral palsy, cardiovascular complications, neurological complications, and mental health conditions. We identified an increased likelihood of prolonged length of stay in almost all cases of these diseases. But the effect of mental health conditions is complicated by the age of the patient and the number of medications administered (during the firs 24 hours of admission) as a proxy for severity of overall illnesses. This indicates that conditions such as pneumonia, sepsis, cerebral palsy, neurological conditions of the brain, abnormalities of heart beat, and complications after surgery may be benefit from targeted interventions tailored to each condition. Interventions relating to mental health conditions such as pervasive development disorders, major depressive disorder, and attention-deficit hyperactivity disorder should consider the age of

the patients as the risk of prolonged length of stay is modified by the age of the patient. In addition, mental health conditions such as conduct disorders (F91), and reaction to severe stress, and adjustment disorders (F43) have risk modified by the overall severity of illness as measured by the number of medications being administered to these patients.

Conditions affecting the respiratory or pulmonary system such as bacterial pneumonia, acute bronchiolitis, acute upper respiratory infections of multiple and unspecified sites (J06), asthma, and respiratory conditions such as pulmonary collapse, disease of bronchus, and disorders of the diaphragm may also be considered for targeted interventions aimed at reducing the probability of unnecessarily prolonged length of stay.

A high model performance was achieved with the mixed effects model with statistics that indicate potential usefulness if implemented electronically with associated interventions. But we considered the large number of statistical interactions discovered in the mixed effects model. These interactions were discovered simply by considering all 2-way statistical interactions between diagnoses and both the age of the child and the number of medications administered during the first 24 hours of hospitalization. This unusually high number of interactions indicated that several other and potentially higher order interactions exist. We explored this possibility bu considering deep extreme gradient boosting trees that explore interactions in computationally feasible ways than can be achieved in equivalent regression models. The resulting machine learning model obtained this way indicated that a statistically significant improvement in model performance was achieved due to the presence of higher order interactions. In other words, the factors driving pediatric hospital length of stay are complex. Unfortunately, unlike regression models, machine learning models do not readily lend themselves to the statistical interpretation of the complexities modelled within them. Even attempts at explanations are obtained by searching for linear approximations. However, pooling all the major findings of this study indicate the need of studies on specific pediatric subpopulations within which specialty-dependent complexities may be modelled

using corresponding mixed effects regressions algorithms. These subpopulations should include the neonatal, mental, and pediatric chronic conditions. These studies are likely to reveal additional insights on the phenomena affecting hospital length of stay that may be helpful in the development of specialty-based intervention protocols.

There were several limitations of this study. The challenges of conducting research with electronic medical records of patients besets this study. The data were not collected for the purpose of research and as such may contain random errors attributable to data entry and storage, as well as differences in the standard of care across organizations. The use of diagnosis codes is particularly nuanced but it is the best data available. Limitations around differences in hospital administration were considered, and their effects mitigated by the choice of mixed effects model that treated hospitals as random intercepts with patients nested within them. Lastly, the implementation of either or both of the statistical and machine learning model would require the use of highly skilled statisticians or data scientists as well as robust information technology team that may be able to integrate statistical and machine learning model in the electronic health records. Notwithstanding, this study identified subgroups within the pediatric population that may be targeted for improvement in quality of care that could result in reduction of unnecessary prolongation of hospitalization. Future studies focusing specifically on the neonatal population is highly encouraged. In addition, studies on pediatric mental health and chronic conditions may result in additional and novel discoveries.

4.4 Summary

Prolonged length of stay (LOS) in medical facilities significantly increases the risk of healthcare-associated infections (HAI) in pediatric and adult patients, often disrupting the access to healthcare and overall health outcomes as well as increasing the financial burden on the entire system. In this study, we designed and implemented a comprehensive study to identify new clinical and demographic factors associated with prolonged length of stay (> 2 weeks) among pediatric patients (aged 18 years and under) in a number of free-standing pediatric and mixed medical facilities. For this, two approaches(univariate and multivariate) were used to conduct statistical analyses. The univariate study helped us explore each variable involved in the final data set, separately. The multivariate study was performed by building a nested mixed effects model. This provided us with valid statistical inferences. The mixed effects model accounts for the hierarchical structure of the data and assesses the statistical significance and effect sizes of age, race/ethnicity, number of medications, medical family history, presence of infection agents (fungi, bacteria,virus), cancer diagnoses and other conditions as well as demographic and clinical variables. Furthermore, a stochastic gradient boosting model was built for prediction to present ways to improve the care quality for patients at high risk of prolonged length of stay, and avoid unnecessary expenses.

Our model identified 11 main effect variables with significant effects on the odds of prolonged length of stay, 12 two-way interaction between age and certain conditions and 33 two-way interaction between the number of medications and certain conditions. The AUC for the mixed effects model was 0.89 and the extreme gradient booster attained AUC of 0.93.

Table 4.2: Summary Statistics LOS Pediatrics

Variables	**Levels**	**LOS 1 week or less n (%) or mean (sd)**	**LOS > 1 week n (%) or mean (sd)**	**p value (chi-squared or t-test)**
Sex	Female	225492 (47.35)	11553 (45.90)	< 0.001
	Male	250737 (52.65)	13618 (54.10)	
Payer	Commercial	152772 (32.08)	6917 (27.48)	< 0.001
	Governmental (Medicare, Medicaid)	184555 (38.75)	11777 (46.79)	
	Other Governmental (Champus, etc)	18226 (3.83)	1001 (3.98)	
	Self-pay	10674 (2.24)	460 (1.83)	
	Other	110002 (23.10)	5016 (19.93)	
Age	AGE	5.93 (6.08)	4.56 (5.97)	< 0.001
Race/Ethnicity	Caucasian	255472 (53.64)	12860 (51.09)	< 0.001
	Hispanic	17889 (3.76)	895 (3.56)	
	Black/African American	106198 (22.30)	5489 (21.81)	
	Asian/Pacific Islander	8702 (1.83)	427 (1.70)	
	Native American	6920 (1.45)	425 (1.69)	
	Other/Unknown	81048 (17.02)	5075 (20.16)	
Previous ED visits (last 6mo)	prevEDcat0	363531 (76.34)	21140 (83.99)	< 0.001
	prevEDcat1	112698 (23.66)	4031 (16.01)	

Maximum previous length of stay (last 6mo)	-	1.52 (4.95)	4.33 (9.59)	< 0.001
Number of medications	Number Of Meds	2.23 (4.45)	11.16 (11.41)	< 0.001
Number of procedures	Number Of Procedure	0.10 (0.51)	0.32 (1.01)	
Emergent Admission	No	77436 (16.26)	3772 (14.99)	< 0.001
	Yes	398793 (83.74)	21399 (85.01)	
Readmission History	No	415569 (87.26)	19766 (78.53)	< 0.001
	Yes	60660 (12.74)	5405 (21.47)	
Free...	is Pediatric 0	236191 (49.60)	10269 (40.80)	< 0.001
	is Pediatric 1	240038 (50.40)	14902 (59.20)	
Viral and other specified intestinal infections (A08)	No	471186 (98.94)	25024 (99.42)	< 0.001
	Yes	5043 (1.06)	147 (0.58)	
Other sepsis (A41)	No	471695 (99.05)	24017 (95.42)	< 0.001
	Yes	4534 (0.95)	1154 (4.58)	
Viral infection of unspecified site (B34)	No	470270 (98.75)	24789 (98.48)	< 0.001
	Yes	5959 (1.25)	382 (1.52)	
Streptococcus, Staphylococcus, and Enterococcus as the cause of diseases classified elsewhere (B95)	No	468183 (98.31)	23936 (95.09)	< 0.001
	Yes	8046 (1.69)	1235 (4.91)	

Other bacterial agents as the cause of diseases classified elsewhere (B96)	No	469698 (98.63)	23888 (94.90)	< 0.001
	Yes	6531 (1.37)	1283 (5.10)	
Viral agents as the cause of diseases classified elsewhere (B97)	No	460317 (96.66)	24431 (97.06)	< 0.001
	Yes	15912 (3.34)	740 (2.94)	
Lymphoid leukemia (C91)	No	470587 (98.82)	24676 (98.03)	< 0.001
	Yes	5642 (1.18)	495 (1.97)	
Sickle-cell disorders (D57)	No	467694 (98.21)	24964 (99.18)	< 0.001
	Yes	8535 (1.79)	207 (0.82)	
Other anemias (D64)	No	466525 (97.96)	23589 (93.71)	< 0.001
	Yes	9704 (2.04)	1582 (6.29)	
Purpura and other hemorrhagic conditions (D69)	No	471560 (99.02)	24293 (96.51)	< 0.001
	Yes	4669 (0.98)	878 (3.49)	
Neutropenia (D70)	No	470668 (98.83)	24190 (96.10)	< 0.001
	Yes	5561 (1.17)	981 (3.90)	
Other disorders of white blood cells (D72)	No	470868 (98.87)	24638 (97.88)	<0.001
	Yes	5361 (1.13)	533 (2.12)	
Type 1 diabetes mellitus (E10)	No	469772 (98.64)	25108 (99.75)	<0.001
	Yes	6457 (1.36)	63 (0.25)	
Overweight and obesity (E66)	No	469530 (98.59)	24848 (98.72)	0.110
	Yes	6699 (1.41)	323 (1.28)	
Volume depletion (E86)	No	449670 (94.42)	24209 (96.18)	< 0.001
	Yes	26559 (5.58)	962 (3.82)	
Other disorders of fluid, electrolyte and acid-base balance (E87)	No	461167 (96.84)	21606 (85.84)	< 0.001
	Yes	15062 (3.16)	3565 (14.16)	

Major depressive disorder, single episode (F32)	No	467416 (98.15)	24435 (97.08)	< 0.001
	Yes	8813 (1.85)	736 (2.92)	
Other anxiety disorders (F41)	No	466936 (98.05)	24511 (97.38)	< 0.001
	Yes	9293 (1.95)	660 (2.62)	
Reaction to severe stress, and adjustment disorders (F43)	No	469188 (98.52)	24576 (97.64)	< 0.001
	Yes	7041 (1.48)	595 (2.36)	
Pervasive developmental disorders (F84)	No	471185 (98.94)	24882 (98.85)	0.190
	Yes	5044 (1.06)	289 (1.15)	
ther disorders of psychological development (F88)	No	471193 (98.94)	24736 (98.27)	<0.001
	Yes	5036 (1.06)	435 (1.73)	
Attention-deficit hyperactivity disorders (F90)	No	464585 (97.55)	24456 (97.16)	<0.001
	Yes	11644 (2.45)	715 (2.84)	
Conduct disorders (F91)	No	471128 (98.93)	24527 (97.44)	< 0.001
	Yes	5101 (1.07)	644 (2.56)	
Epilepsy and recurrent seizures (G40)	No	458341 (96.24)	24071 (95.63)	< 0.001
	Yes	17888 (3.76)	1100 (4.37)	
Sleep disorders (G47)	No	467634 (98.20)	24355 (96.76)	< 0.001
	Yes	8595 (1.80)	816 (3.24)	
Cerebral palsy (G80)	No	469797 (98.65)	24653 (97.94)	< 0.001
	Yes	6432 (1.35)	518 (2.06)	
Other disorders of brain (G93)	No	468274 (98.33)	23958 (95.18)	< 0.001
	Yes	7955 (1.67)	1213 (4.82)	

Suppurative and unspecified otitis media (H66)	No	468492 (98.38)	25038 (99.47)	< 0.001
	Yes	7737 (1.62)	133 (0.53)	
Essential (primary) hypertension (I10)	No	471850 (99.08)	24404 (96.95)	< 0.001
	Yes	4379 (0.92)	767 (3.05)	
Other cardiac arrhythmias (I49)	No	471856 (99.08)	24354 (96.75)	< 0.001
	Yes	4373 (0.92)	817 (3.25)	
Acute upper respiratory infections of multiple and unspecified sites (J06)	No	463782 (97.39)	24827 (98.63)	< 0.001
	Yes	12447 (2.61)	344 (1.37)	
Bacterial pneumonia, not elsewhere classified (J15)	No	470220 (98.74)	24465 (97.20)	< 0.001
	Yes	6009 (1.26)	706 (2.80)	
Pneumonia, unspecified organism (J18)	No	459014 (96.39)	24188 (96.09)	0.017
	Yes	17215 (3.61)	983 (3.91)	
Acute bronchiolitis (J21)	No	455023 (95.55)	24546 (97.52)	<0.001
	Yes	21206 (4.45)	625 (2.48)	
Asthma (J45)	No	432614 (90.84)	24200 (96.14)	<0.001
	Yes	43615 (9.16)	971 (3.86)	
Acute respiratory distress syndrome (J80)	No	470722 (98.84)	24815 (98.59)	<0.001
	Yes	5507 (1.16)	356 (1.41)	
Respiratory failure, not elsewhere classified (J96)	No	467134 (98.09)	22885 (90.92)	< 0.001
	Yes	9095 (1.91)	2286 (9.08)	
Other respiratory disorders (J98)	No	465501 (97.75)	23391 (92.93)	< 0.001
	Yes	10728 (2.25)	1780 (7.07)	
Gastro-esophageal reflux disease K21	No	460106 (96.61)	23263 (92.42)	< 0.001
	Yes	16123 (3.39)	1908 (7.58)	
Acute appendicitis (K35)	No	468467 (98.37)	25016 (99.38)	< 0.001
	Yes	7762 (1.63)	155 (0.62)	

Other and unspecified noninfective gastroenteritis and colitis (K52)	No	466229 (97.90)	24407 (96.96)	< 0.001
	Yes	10000 (2.10)	764 (3.04)	
Paralytic ileus and intestinal obstruction without hernia (K56)	No	471611 (99.03)	24491 (97.30)	<0.001
	Yes	4618 (0.97)	680 (2.70)	
Other functional intestinal disorders (K59)	No	457981 (96.17)	23391 (92.93)	<0.001
	Yes	18248 (3.83)	1780 (7.07)	
Cellulitis and acute lymphangitis (L03)	No	465002 (97.64)	24720 (98.21)	<0.001
	Yes	11227 (2.36)	451 (1.79)	
Scoliosis (M41)	No	471283 (98.96)	24868 (98.80)	0.013
	Yes	4946 (1.04)	303 (1.20)	
Other and unspecified soft tissue disorders, not elsewhere classified (M79)	No	471580 (99.02)	24875 (98.82)	0.002
	Yes	4649 (0.98)	296 (1.18)	
Obstructive and reflux uropathy (N13)	No	470817 (98.86)	24689 (98.09)	< 0.001
	Yes	5412 (1.14)	482 (1.91)	
Other disorders of urinary system (N39)	No	468334 (98.34)	24484 (97.27)	< 0.001
	Yes	7895 (1.66)	687 (2.73)	
Disorders of newborn related to short gestation and low birth weight, not elsewhere classified (P07)	No	469040 (98.49)	20310 (80.69)	< 0.001
	Yes	7189 (1.51)	4861 (19.31)	
Disorders of newborn related to long gestation and high birth weight (P08)	No	469849 (98.66)	24934 (99.06)	< 0.001
	Yes	6380 (1.34)	237 (0.94)	
Respiratory distress of newborn (P22)	No	471406 (98.99)	22051 (87.60)	< 0.001

	Yes	4823 (1.01)	3120 (12.40)	
Other respiratory conditions originating in the perinatal period (P28)	No	470980 (98.90)	21584 (85.75)	< 0.001
	Yes	5249 (1.10)	3587 (14.25)	
Cardiovascular disorders originating in the perinatal period (P29)	No	471319 (98.97)	22500 (89.39)	< 0.001
	Yes	4910 (1.03)	2671 (10.61)	
Neonatal jaundice from other and unspecified causes (P59)	No	460905 (96.78)	21370 (84.90)	< 0.001
	Yes	15324 (3.22)	3801 (15.10)	
Transitory disorders of carbohydrate metabolism specific to newborn (P70)	No	472203 (99.15)	23975 (95.25)	< 0.001
	Yes	4026 (0.85)	1196 (4.75)	
Feeding problems of newborn (P92)	No	471859 (99.08)	22596 (89.77)	< 0.001
	Yes	4370 (0.92)	2575 (10.23)	
Other conditions originating in the perinatal period (P96)	No	470078 (98.71)	23245 (92.35)	< 0.001
	Yes	6151 (1.29)	1926 (7.65)	
Congenital malformations of cardiac septa (Q21)	No	467151 (98.09)	22430 (89.11)	< 0.001
	Yes	9078 (1.91)	2741 (10.89)	
Congenital malformations of great arteries (Q25)	No	472180 (99.15)	23090 (91.73)	< 0.001
	Yes	4049 (0.85)	2081 (8.27)	
Abnormalities of heart beat (R00)	No	462313 (97.08)	23585 (93.70)	< 0.001
	Yes	13916 (2.92)	1586 (6.30)	
Cough (R05)	No	469262 (98.54)	24923 (99.01)	< 0.001
	Yes	6967 (1.46)	248 (0.99)	
Abnormalities of breathing (R06)	No	446092 (93.67)	22427 (89.10)	<0.001
	Yes	30137 (6.33)	2744 (10.90)	

Other symptoms and signs involving the circulatory and respiratory system (R09)	No	454487 (95.43)	24095 (95.73)	0.032
	Yes	21742 (4.57)	1076 (4.27)	
Abdominal and pelvic pain (R10)	No	460785 (96.76)	24550 (97.53)	<0.001
	Yes	15444 (3.24)	621 (2.47)	
Nausea and vomiting (R11)	No	454020 (95.34)	23773 (94.45)	<0.001
	Yes	22209 (4.66)	1398 (5.55)	
Other symptoms and signs involving the digestive system and abdomen (R19)	No	467938 (98.26)	24396 (96.92)	<0.001
	Yes	8291 (1.74)	775 (3.08)	
Symptoms and signs involving emotional state (R45)	No	468191 (98.31)	24591 (97.70)	<0.001
	Yes	8038 (1.69)	580 (2.30)	
Fever of other and unknown origin (R50)	No	442745 (92.97)	23372 (92.85)	0.491
	Yes	33484 (7.03)	1799 (7.15)	
Headache (R51)	No	470066 (98.71)	24819 (98.60)	0.163
	Yes	6163 (1.29)	352 (1.40)	
Convulsions, not elsewhere classified (R56)	No	462717 (97.16)	24526 (97.44)	0.011
	Yes	13512 (2.84)	645 (2.56)	
Lack of expected normal physiological development in childhood and adults (R62)	No	462776 (97.18)	23701 (94.16)	< 0.001
	Yes	13453 (2.82)	1470 (5.84)	
Symptoms and signs concerning food and fluid intake (R63)	No	462577 (97.13)	22677 (90.09)	< 0.001
	Yes	13652 (2.87)	2494 (9.91)	
Findings of drugs and other substances, not normally found in blood (R78)	No	471642 (99.04)	24431 (97.06)	< 0.001
	Yes	4587 (0.96)	740 (2.94)	

Surgical operation and other surgical procedures as the cause of abnormal reaction of the patient, or of later complication, without mention of misadventure at the time of the procedure (Y83)	No	471665 (99.04)	24491 (97.30)	< 0.001
	Yes	4564 (0.96)	680 (2.70)	
Place of occurrence of the external cause (Y92)	No	463240 (97.27)	24308 (96.57)	< 0.001
	Yes	12989 (2.73)	863 (3.43)	
Encounter for observation and evaluation of newborn for suspected diseases and conditions ruled out (Z05)	No	469243 (98.53)	23633 (93.89)	<0.001
	Yes	6986 (1.47)	1538 (6.11)	
Encounter for immunization (Z23)	No	459579 (96.50)	23633 (93.89)	<0.001
	Yes	16650 (3.50)	1538 (6.11)	
Outcome of delivery (Z37)	No	468046 (98.28)	25032 (99.45)	<0.001
	Yes	8183 (1.72)	139 (0.55)	
Liveborn infants according to place of birth and type of delivery (Z38)	No	418130 (87.80)	22153 (88.01)	0.326
	Yes	58099 (12.20)	3018 (11.99)	
Encounter for other aftercare and medical care (Z51)	No	462358 (97.09)	23759 (94.39)	<0.001
	Yes	13871 (2.91)	1412 (5.61)	
Persons encountering health services for other counseling and medical advice, not elsewhere classified (Z71)	No	470703 (98.84)	24641 (97.89)	<0.001
	Yes	5526 (1.16)	530 (2.11)	

Long term (current) drug therapy (Z79)	No	468219 (98.32)	24619 (97.81)	<0.001
	Yes	8010 (1.68)	552 (2.19)	
Family history of certain disabilities and chronic diseases (leading to disablement) (Z82)	No	467499 (98.17)	24836 (98.67)	<0.001
	Yes	8730 (1.83)	335 (1.33)	
Family history of other specific disorders (Z83)	No	470836 (98.87)	24896 (98.91)	0.580
	Yes	5393 (1.13)	275 (1.09)	
Personal history of certain other diseases (Z86)	No	470511 (98.80)	24675 (98.03)	<0.001
	Yes	5718 (1.20)	496 (1.97)	
Personal history of other diseases and conditions (Z87)	No	466394 (97.93)	24455 (97.16)	<0.001
	Yes	9835 (2.07)	716 (2.84)	
Personal risk factors, not elsewhere classified (Z91)	No	464485 (97.53)	24473 (97.23)	0.002
	Yes	11744 (2.47)	698 (2.77)	
Artificial opening status (Z93)	No	465233 (97.69)	23840 (94.71)	<0.001
	Yes	10996 (2.31)	1331 (5.29)	
Other postprocedural states (Z98)	No	466131 (97.88)	24328 (96.65)	<0.001
	Yes	10098 (2.12)	843 (3.35)	

Table 4.3: Results of the multivariate statistical analysis

Variable	**Levels**	**Odds ratio**	**p value**
Sex	Female	Reference	< 0.001
	Male	0.943 (0.914, 0.973)	
Payer	Commercial	Reference	
	Governmental (Medicare, Medicaid)	1.347 (1.293, 1.404)	< 0.001
	Other Governmental (Champus, etc)	1.212 (1.114, 1.319)	
	Self-pay	1.296 (1.157, 1.452)	
	Other	1.343 (1.275, 1.413)	
Race/Ethnicity	Caucasian	Reference	
	Hispanic	1.106 (1.014, 1.207)	0.023
	Black/African American	1.140 (1.091, 1.191)	< 0.001
	Asian/Pacific Islander	1.004 (0.893, 1.129)	0.947
	Native American	1.145 (0.995, 1.317)	0.059
	Other/Unknown	1.112 (1.061, 1.166)	< 0.001

Previous ED visits (last 6mo)	-	0.621 (0.595, 0.649)	< 0.001
Maximum previous length of stay (last 6mo)	-	1.042 (1.040, 1.044)	< 0.001
Free-standing pediatric hospital	Yes	1.431 (0.659, 3.107)	0.364
EmergentAdmission	Yes	1.672 (1.598, 1.749)	< 0.001
ReadmissionHistory	Yes	1.179 (1.124, 1.236)	< 0.001
Other sepsis (A41)	Yes	1.599 (1.449, 1.764)	< 0.001
Other bacterial agents as the cause of diseases classified elsewhere (B96)	Yes	1.912 (1.740, 2.100)	< 0.001
Lymphoid leukemia (C91)	Yes	0.562 (0.494, 0.640)	< 0.001
Type 1 diabetes mellitus (E10)	Yes	0.279 (0.211, 0.369)	< 0.001
Cerebral palsy (G80)	Yes	1.314 (1.156, 1.494)	< 0.001
Other disorders of brain (G93)	Yes	1.224 (1.119, 1.340)	< 0.001
Suppurative and unspecified otitis media (H66)	Yes	0.522 (0.422, 0.645)	< 0.001
Bacterial pneumonia, not elsewhere classified (J15)	Yes	1.867 (1.664, 2.094)	< 0.001
Pneumonia, unspecified organism (J18)	Yes	0.796 (0.727, 0.872)	< 0.001
Disorders of newborn related to short gestation and low birth weight, not elsewhere classified (P07)	Yes	5.466 (5.084, 5.877)	< 0.001
Respiratory distress of newborn (P22)	Yes	1.839 (1.687, 2.004)	< 0.001
Other respiratory conditions originating in the perinatal period (P28)	Yes	2.579 (2.368, 2.810)	< 0.001
Cardiovascular disorders originating in the perinatal period (P29)	Yes	1.184 (1.071, 1.309)	< 0.001

Other conditions originating in the perinatal period (P96)	Yes	2.412 (2.180, 2.668)	< 0.001
Abnormalities of heart beat (R00)	Yes	1.106 (1.024, 1.195)	0.010
Symptoms and signs concerning food and fluid intake (R63)	Yes	1.630 (1.525, 1.742)	< 0.001
Surgical operation and other surgical procedures as the cause of abnormal reaction of the patient, or of later complication, without mention of misadventure at the time of the procedure (Y83)	Yes	1.227 (1.091, 1.380)	< 0.001
Encounter for observation and evaluation of newborn for suspected diseases and conditions ruled out (Z05)	Yes	1.239 (1.126, 1.363)	< 0.001
Encounter for immunization (Z23)	Yes	1.133 (1.029, 1.248)	0.011
Family history of certain disabilities and chronic diseases (leading to disablement) (Z82)	Yes	0.641 (0.551, 0.745)	< 0.001
Personal history of other diseases and conditions (Z87)	Yes	0.816 (0.734, 0.907)	< 0.001
Other postprocedural states (Z98)	Yes	0.698 (0.632, 0.770)	< 0.001
Pervasive developmental disorders (F84)	Yes	0.577 (0.373, 0.893)	0.013
Sleep disorders (G47)	Yes	0.989 (0.842, 1.162)	0.891
Other functional intestinal disorders (K59)	Yes	1.028 (0.895, 1.181)	0.699
Reaction to severe stress, and adjustment disorders (F43)	Yes	1.354 (1.157, 1.585)	< 0.001

Streptococcus, Staphylococcus, and Enterococcus as the cause of diseases classified elsewhere (B95)	Yes	1.031 (1.017, 1.046)	< 0.001
Other anemias (D64)	Yes	0.965 (0.953, 0.978)	< 0.001
Major depressive disorder, single episode (F32)	Yes	0.946 (0.918, 0.976)	< 0.001
Pervasive developmental disorders (F84)	Yes	1.063 (1.026, 1.102)	< 0.001
Attention-deficit hyperactivity disorders (F90)	Yes	0.934 (0.909, 0.960)	< 0.001
Sleep disorders (G47)	Yes	1.031 (1.014, 1.047)	< 0.001
Asthma (J45)	Yes	1.052 (1.036, 1.068)	< 0.001
Other functional intestinal disorders (K59)	Yes	1.032 (1.020, 1.044)	< 0.001
Obstructive and reflux uropathy (N13)	Yes	0.947 (0.921, 0.974)	< 0.001
Other disorders of urinary system (N39)	Yes	1.034 (1.016, 1.052)	< 0.001
Convulsions, not elsewhere classified (R56)	Yes	0.956 (0.936, 0.977)	< 0.001
Outcome of delivery (Z37)	Yes	0.904 (0.873, 0.935)	< 0.001
Viral and other specified intestinal infections (A08)	Present	1.096 (1.064, 1.128)	< 0.001
Viral agents as the cause of diseases classified elsewhere (B97)	Present	1.044 (1.031, 1.058)	< 0.001
Sickle-cell disorders (D57)	Present	1.042 (1.023, 1.062)	< 0.001
Neutropenia (D70)	Present	1.035 (1.023, 1.047)	< 0.001
Volume depletion (E86)	Present	1.048 (1.038, 1.058)	< 0.001

Other disorders of fluid, electrolyte and acid-base balance (E87)	Present	0.980 (0.975, 0.986)	< 0.001
Reaction to severe stress, and adjustment disorders (F43)	Present	1.032 (1.018, 1.047)	< 0.001
Conduct disorders (F91)	Present	1.066 (1.044, 1.089)	< 0.001
Epilepsy and recurrent seizures (G40)	Present	1.026 (1.017, 1.035)	< 0.001
Essential (primary) hypertension (I10)	Present	0.969 (0.959, 0.978)	< 0.001
Acute upper respiratory infections of multiple and unspecified sites (J06)	Present	1.047 (1.029, 1.065)	< 0.001
Acute bronchiolitis (J21)	Present	1.086 (1.071, 1.101)	< 0.001
Asthma (J45)	Present	1.026 (1.017, 1.034)	< 0.001
Respiratory failure, not elsewhere classified (J96)	Present	0.977 (0.970, 0.983)	< 0.001
Other respiratory disorders (J98)	Present	0.986 (0.979, 0.994)	< 0.001
Gastro-esophageal reflux disease (K21)	Present	1.025 (1.016, 1.033)	< 0.001
Acute appendicitis (K35)	Present	1.076 (1.052, 1.100)	< 0.001
Other functional intestinal disorders (K59)	Present	0.987 (0.980, 0.994)	< 0.001
Cellulitis and acute lymphangitis (L03)	Present	1.065 (1.050, 1.080)	< 0.001
Scoliosis (M41)	Present	0.964 (0.951, 0.976)	< 0.001
Disorders of newborn related to long gestation and high birth weight (P08)	Present	1.124 (1.080, 1.169)	< 0.001
Neonatal jaundice from other and unspecified causes (P59)	Present	1.049 (1.035, 1.063)	< 0.001
Transitory disorders of carbohydrate metabolism specific to newborn (P70)	Present	1.044 (1.022, 1.067)	< 0.001
Feeding problems of newborn (P92)	Present	1.076 (1.059, 1.093)	< 0.001

Congenital malformations of cardiac septa (Q21)	Present	0.955 (0.949, 0.962)	< 0.001
Congenital malformations of great arteries (Q25)	Present	0.965 (0.957, 0.974)	< 0.001
Abnormalities of breathing (R06)	Present	1.043 (1.036, 1.051)	< 0.001
Other symptoms and signs involving the circulatory and respiratory system (R09)	Present	1.022 (1.013, 1.032)	< 0.001
Abdominal and pelvic pain (R10)	Present	1.035 (1.023, 1.047)	< 0.001
Nausea and vomiting (R11)	Present	1.028 (1.019, 1.036)	< 0.001
Symptoms and signs involving emotional state (R45)	Present	1.052 (1.035, 1.069)	< 0.001
Fever of other and unknown origin (R50)	Present	1.036 (1.028, 1.043)	< 0.001
Convulsions, not elsewhere classified (R56)	Present	1.053 (1.041, 1.065)	< 0.001
Lack of expected normal physiological development in childhood and adults (R62)	Present	1.023 (1.015, 1.032)	< 0.001
Place of occurrence of the external cause (Y92)	Present	1.034 (1.023, 1.044)	< 0.001
Outcome of delivery (Z37)	Present	1.047 (1.019, 1.076)	< 0.001
Liveborn infants according to place of birth and type of delivery (Z38)	Present	1.369 (1.342, 1.395)	< 0.001
Encounter for other aftercare and medical care (Z51)	Present	1.048 (1.039, 1.057)	< 0.001

Personal risk factors, not elsewhere classified (Z91)	Present	1.024 (1.012, 1.036)	< 0.001
Artificial opening status (Z93)	Present	1.024 (1.014, 1.033)	< 0.001

Table 4.5: Summary Statistics LOS Pediatrics

Variables	**Levels**	**LOS 1 week or less n (%) or mean (sd)**	**LOS > 1 week n (%) or mean (sd)**	**p value (chi-squared or t-test)**
Sex	Female	225492 (47.35)	11553 (45.90)	< 0.001
	Male	250737 (52.65)	13618 (54.10)	
Payer	Commercial	152772 (32.08)	6917 (27.48)	< 0.001
	Governmental (Medicare, Medicaid)	184555 (38.75)	11777 (46.79)	
	Other Governmental (Champus, etc)	18226 (3.83)	1001 (3.98)	
	Self-pay	10674 (2.24)	460 (1.83)	
	Other	110002 (23.10)	5016 (19.93)	
Age	AGE	5.93 (6.08)	4.56 (5.97)	< 0.001

Race/Ethnicity	Caucasian	255472 (53.64)	12860 (51.09)	< 0.001
	Hispanic	17889 (3.76)	895 (3.56)	
	Black/African American	106198 (22.30)	5489 (21.81)	
	Asian/Pacific Islander	8702 (1.83)	427 (1.70)	
	Native American	6920 (1.45)	425 (1.69)	
	Other/Unknown	81048 (17.02)	5075 (20.16)	
Previous ED visits (last 6mo)	prevEDcat0	363531 (76.34)	21140 (83.99)	< 0.001
	prevEDcat1	112698 (23.66)	4031 (16.01)	
Maximum previous length of stay (last 6mo)	-	1.52 (4.95)	4.33 (9.59)	< 0.001
Number of medications	Number Of Meds	2.23 (4.45)	11.16 (11.41)	< 0.001
Number of procedures	Number Of Procedure	0.10 (0.51)	0.32 (1.01)	
Emergent Admission	No	77436 (16.26)	3772 (14.99)	< 0.001
	Yes	398793 (83.74)	21399 (85.01)	
Readmission History	No	415569 (87.26)	19766 (78.53)	< 0.001
	Yes	60660 (12.74)	5405 (21.47)	
Free...	is Pediatric 0	236191 (49.60)	10269 (40.80)	< 0.001

	is Pediatric 1	240038 (50.40)	14902 (59.20)	
Viral and other specified intestinal infections (A08)	No	471186 (98.94)	25024 (99.42)	< 0.001
	Yes	5043 (1.06)	147 (0.58)	
Other sepsis (A41)	No	471695 (99.05)	24017 (95.42)	< 0.001
	Yes	4534 (0.95)	1154 (4.58)	
Viral infection of unspecified site (B34)	No	470270 (98.75)	24789 (98.48)	< 0.001
	Yes	5959 (1.25)	382 (1.52)	
Streptococcus, Staphylococcus, and Enterococcus as the cause of diseases classified elsewhere (B95)	No	468183 (98.31)	23936 (95.09)	< 0.001
	Yes	8046 (1.69)	1235 (4.91)	
Other bacterial agents as the cause of diseases classified elsewhere (B96)	No	469698 (98.63)	23888 (94.90)	< 0.001
	Yes	6531 (1.37)	1283 (5.10)	
Viral agents as the cause of diseases classified elsewhere (B97)	No	460317 (96.66)	24431 (97.06)	< 0.001
	Yes	15912 (3.34)	740 (2.94)	
Lymphoid leukemia (C91)	No	470587 (98.82)	24676 (98.03)	< 0.001
	Yes	5642 (1.18)	495 (1.97)	
Sickle-cell disorders (D57)	No	467694 (98.21)	24964 (99.18)	< 0.001

	Yes	8535 (1.79)	207 (0.82)	
Other anemias (D64)	No	466525 (97.96)	23589 (93.71)	< 0.001
	Yes	9704 (2.04)	1582 (6.29)	
Purpura and other hemorrhagic conditions (D69)	No	471560 (99.02)	24293 (96.51)	< 0.001
	Yes	4669 (0.98)	878 (3.49)	
Neutropenia (D70)	No	470668 (98.83)	24190 (96.10)	< 0.001
	Yes	5561 (1.17)	981 (3.90)	
Other disorders of white blood cells (D72)	No	470868 (98.87)	24638 (97.88)	<0.001
	Yes	5361 (1.13)	533 (2.12)	
Type 1 diabetes mellitus (E10)	No	469772 (98.64)	25108 (99.75)	<0.001
	Yes	6457 (1.36)	63 (0.25)	
Overweight and obesity (E66)	No	469530 (98.59)	24848 (98.72)	0.110
	Yes	6699 (1.41)	323 (1.28)	
Volume depletion (E86)	No	449670 (94.42)	24209 (96.18)	< 0.001
	Yes	26559 (5.58)	962 (3.82)	
Other disorders of fluid, electrolyte and acid-base balance (E87)	No	461167 (96.84)	21606 (85.84)	< 0.001
	Yes	15062 (3.16)	3565 (14.16)	
Major depressive disorder, single episode (F32)	No	467416 (98.15)	24435 (97.08)	< 0.001

	Yes	8813 (1.85)	736 (2.92)	
Other anxiety disorders (F41)	No	466936 (98.05)	24511 (97.38)	< 0.001
	Yes	9293 (1.95)	660 (2.62)	
Reaction to severe stress, and adjustment disorders (F43)	No	469188 (98.52)	24576 (97.64)	< 0.001
	Yes	7041 (1.48)	595 (2.36)	
Pervasive developmental disorders (F84)	No	471185 (98.94)	24882 (98.85)	0.190
	Yes	5044 (1.06)	289 (1.15)	
ther disorders of psychological development (F88)	No	471193 (98.94)	24736 (98.27)	<0.001
	Yes	5036 (1.06)	435 (1.73)	
Attention-deficit hyperactivity disorders (F90)	No	464585 (97.55)	24456 (97.16)	<0.001
	Yes	11644 (2.45)	715 (2.84)	
Conduct disorders (F91)	No	471128 (98.93)	24527 (97.44)	< 0.001
	Yes	5101 (1.07)	644 (2.56)	
Epilepsy and recurrent seizures (G40)	No	458341 (96.24)	24071 (95.63)	< 0.001
	Yes	17888 (3.76)	1100 (4.37)	
Sleep disorders (G47)	No	467634 (98.20)	24355 (96.76)	< 0.001
	Yes	8595 (1.80)	816 (3.24)	
Cerebral palsy (G80)	No	469797 (98.65)	24653 (97.94)	< 0.001

	Yes	6432 (1.35)	518 (2.06)	
Other disorders of brain (G93)	No	468274 (98.33)	23958 (95.18)	< 0.001
	Yes	7955 (1.67)	1213 (4.82)	
Suppurative and unspecified otitis media (H66)	No	468492 (98.38)	25038 (99.47)	< 0.001
	Yes	7737 (1.62)	133 (0.53)	
Essential (primary) hypertension (I10)	No	471850 (99.08)	24404 (96.95)	< 0.001
	Yes	4379 (0.92)	767 (3.05)	
Other cardiac arrhythmias (I49)	No	471856 (99.08)	24354 (96.75)	< 0.001
	Yes	4373 (0.92)	817 (3.25)	
Acute upper respiratory infections of multiple and unspecified sites (J06)	No	463782 (97.39)	24827 (98.63)	< 0.001
	Yes	12447 (2.61)	344 (1.37)	
Bacterial pneumonia, not elsewhere classified (J15)	No	470220 (98.74)	24465 (97.20)	< 0.001
	Yes	6009 (1.26)	706 (2.80)	
Pneumonia, unspecified organism (J18)	No	459014 (96.39)	24188 (96.09)	0.017
	Yes	17215 (3.61)	983 (3.91)	
Acute bronchiolitis (J21)	No	455023 (95.55)	24546 (97.52)	<0.001
	Yes	21206 (4.45)	625 (2.48)	
Asthma (J45)	No	432614 (90.84)	24200 (96.14)	<0.001

	Yes	43615 (9.16)	971 (3.86)	
Acute respiratory distress syndrome (J80)	No	470722 (98.84)	24815 (98.59)	<0.001
	Yes	5507 (1.16)	356 (1.41)	
Respiratory failure, not elsewhere classified (J96)	No	467134 (98.09)	22885 (90.92)	< 0.001
	Yes	9095 (1.91)	2286 (9.08)	
Other respiratory disorders (J98)	No	465501 (97.75)	23391 (92.93)	< 0.001
	Yes	10728 (2.25)	1780 (7.07)	
Gastro-esophageal reflux disease K21	No	460106 (96.61)	23263 (92.42)	< 0.001
	Yes	16123 (3.39)	1908 (7.58)	
Acute appendicitis (K35)	No	468467 (98.37)	25016 (99.38)	< 0.001
	Yes	7762 (1.63)	155 (0.62)	
Other and unspecified noninfective gastroenteritis and colitis (K52)	No	466229 (97.90)	24407 (96.96)	< 0.001
	Yes	10000 (2.10)	764 (3.04)	
Paralytic ileus and intestinal obstruction without hernia (K56)	No	471611 (99.03)	24491 (97.30)	<0.001
	Yes	4618 (0.97)	680 (2.70)	
Other functional intestinal disorders (K59)	No	457981 (96.17)	23391 (92.93)	<0.001
	Yes	18248 (3.83)	1780 (7.07)	
Cellulitis and acute lymphangitis (L03)	No	465002 (97.64)	24720 (98.21)	<0.001

	Yes	11227 (2.36)	451 (1.79)	
Scoliosis (M41)	No	471283 (98.96)	24868 (98.80)	0.013
	Yes	4946 (1.04)	303 (1.20)	
Other and unspecified soft tissue disorders, not elsewhere classified (M79)	No	471580 (99.02)	24875 (98.82)	0.002
	Yes	4649 (0.98)	296 (1.18)	
Obstructive and reflux uropathy (N13)	No	470817 (98.86)	24689 (98.09)	< 0.001
	Yes	5412 (1.14)	482 (1.91)	
Other disorders of urinary system (N39)	No	468334 (98.34)	24484 (97.27)	< 0.001
	Yes	7895 (1.66)	687 (2.73)	
Disorders of newborn related to short gestation and low birth weight, not elsewhere classified (P07)	No	469040 (98.49)	20310 (80.69)	< 0.001
	Yes	7189 (1.51)	4861 (19.31)	
Disorders of newborn related to long gestation and high birth weight (P08)	No	469849 (98.66)	24934 (99.06)	< 0.001
	Yes	6380 (1.34)	237 (0.94)	
Respiratory distress of newborn (P22)	No	471406 (98.99)	22051 (87.60)	< 0.001
	Yes	4823 (1.01)	3120 (12.40)	
Other respiratory conditions originating in the perinatal period (P28)	No	470980 (98.90)	21584 (85.75)	< 0.001
	Yes	5249 (1.10)	3587 (14.25)	

Cardiovascular disorders originating in the perinatal period (P29)	No	471319 (98.97)	22500 (89.39)	< 0.001
	Yes	4910 (1.03)	2671 (10.61)	
Neonatal jaundice from other and unspecified causes (P59)	No	460905 (96.78)	21370 (84.90)	< 0.001
	Yes	15324 (3.22)	3801 (15.10)	
Transitory disorders of carbohydrate metabolism specific to newborn (P70)	No	472203 (99.15)	23975 (95.25)	< 0.001
	Yes	4026 (0.85)	1196 (4.75)	
Feeding problems of newborn (P92)	No	471859 (99.08)	22596 (89.77)	< 0.001
	Yes	4370 (0.92)	2575 (10.23)	
Other conditions originating in the perinatal period (P96)	No	470078 (98.71)	23245 (92.35)	< 0.001
	Yes	6151 (1.29)	1926 (7.65)	
Congenital malformations of cardiac septa (Q21)	No	467151 (98.09)	22430 (89.11)	< 0.001
	Yes	9078 (1.91)	2741 (10.89)	
Congenital malformations of great arteries (Q25)	No	472180 (99.15)	23090 (91.73)	< 0.001
	Yes	4049 (0.85)	2081 (8.27)	
Abnormalities of heart beat (R00)	No	462313 (97.08)	23585 (93.70)	< 0.001
	Yes	13916 (2.92)	1586 (6.30)	
Cough (R05)	No	469262 (98.54)	24923 (99.01)	< 0.001
	Yes	6967 (1.46)	248 (0.99)	

Abnormalities of breathing (R06)	No	446092 (93.67)	22427 (89.10)	<0.001
	Yes	30137 (6.33)	2744 (10.90)	
Other symptoms and signs involving the circulatory and respiratory system (R09)	No	454487 (95.43)	24095 (95.73)	0.032
	Yes	21742 (4.57)	1076 (4.27)	
Abdominal and pelvic pain (R10)	No	460785 (96.76)	24550 (97.53)	<0.001
	Yes	15444 (3.24)	621 (2.47)	
Nausea and vomiting (R11)	No	454020 (95.34)	23773 (94.45)	<0.001
	Yes	22209 (4.66)	1398 (5.55)	
Other symptoms and signs involving the digestive system and abdomen (R19)	No	467938 (98.26)	24396 (96.92)	<0.001
	Yes	8291 (1.74)	775 (3.08)	
Symptoms and signs involving emotional state (R45)	No	468191 (98.31)	24591 (97.70)	<0.001
	Yes	8038 (1.69)	580 (2.30)	
Fever of other and unknown origin (R50)	No	442745 (92.97)	23372 (92.85)	0.491
	Yes	33484 (7.03)	1799 (7.15)	
Headache (R51)	No	470066 (98.71)	24819 (98.60)	0.163
	Yes	6163 (1.29)	352 (1.40)	
Convulsions, not elsewhere classified (R56)	No	462717 (97.16)	24526 (97.44)	0.011
	Yes	13512 (2.84)	645 (2.56)	

Lack of expected normal physiological development in childhood and adults (R62)	No	462776 (97.18)	23701 (94.16)	< 0.001
	Yes	13453 (2.82)	1470 (5.84)	
Symptoms and signs concerning food and fluid intake (R63)	No	462577 (97.13)	22677 (90.09)	< 0.001
	Yes	13652 (2.87)	2494 (9.91)	
Findings of drugs and other substances, not normally found in blood (R78)	No	471642 (99.04)	24431 (97.06)	< 0.001
	Yes	4587 (0.96)	740 (2.94)	
Surgical operation and other surgical procedures as the cause of abnormal reaction of the patient, or of later complication, without mention of misadventure at the time of the procedure (Y83)	No	471665 (99.04)	24491 (97.30)	< 0.001
	Yes	4564 (0.96)	680 (2.70)	
Place of occurrence of the external cause (Y92)	No	463240 (97.27)	24308 (96.57)	< 0.001
	Yes	12989 (2.73)	863 (3.43)	
Encounter for observation and evaluation of newborn for suspected diseases and conditions ruled out (Z05)	No	469243 (98.53)	23633 (93.89)	<0.001
	Yes	6986 (1.47)	1538 (6.11)	
Encounter for immunization (Z23)	No	459579 (96.50)	23633 (93.89)	<0.001
	Yes	16650 (3.50)	1538 (6.11)	

Outcome of delivery (Z37)	No	468046 (98.28)	25032 (99.45)	<0.001
	Yes	8183 (1.72)	139 (0.55)	
Liveborn infants according to place of birth and type of delivery (Z38)	No	418130 (87.80)	22153 (88.01)	0.326
	Yes	58099 (12.20)	3018 (11.99)	
Encounter for other aftercare and medical care (Z51)	No	462358 (97.09)	23759 (94.39)	<0.001
	Yes	13871 (2.91)	1412 (5.61)	
Persons encountering health services for other counseling and medical advice, not elsewhere classified (Z71)	No	470703 (98.84)	24641 (97.89)	<0.001
	Yes	5526 (1.16)	530 (2.11)	
Long term (current) drug therapy (Z79)	No	468219 (98.32)	24619 (97.81)	<0.001
	Yes	8010 (1.68)	552 (2.19)	
Family history of certain disabilities and chronic diseases (leading to disablement) (Z82)	No	467499 (98.17)	24836 (98.67)	<0.001
	Yes	8730 (1.83)	335 (1.33)	
Family history of other specific disorders (Z83)	No	470836 (98.87)	24896 (98.91)	0.580
	Yes	5393 (1.13)	275 (1.09)	
Personal history of certain other diseases (Z86)	No	470511 (98.80)	24675 (98.03)	<0.001
	Yes	5718 (1.20)	496 (1.97)	

Personal history of other diseases and conditions (Z87)	No	466394 (97.93)	24455 (97.16)	<0.001
	Yes	9835 (2.07)	716 (2.84)	
Personal risk factors, not elsewhere classified (Z91)	No	464485 (97.53)	24473 (97.23)	0.002
	Yes	11744 (2.47)	698 (2.77)	
Artificial opening status (Z93)	No	465233 (97.69)	23840 (94.71)	<0.001
	Yes	10996 (2.31)	1331 (5.29)	
Other postprocedural states (Z98)	No	466131 (97.88)	24328 (96.65)	<0.001
	Yes	10098 (2.12)	843 (3.35)	

Table 4.6: Results of the multivariate statistical analysis

Variable	**Levels**	**Odds ratio**	**p value**
Sex	Female	Reference	< 0.001
	Male	0.943 (0.914, 0.973)	
Payer	Commercial	Reference	
	Governmental (Medicare, Medicaid)	1.347 (1.293, 1.404)	< 0.001
	Other Governmental (Champus, etc)	1.212 (1.114, 1.319)	
	Self-pay	1.296 (1.157, 1.452)	
	Other	1.343 (1.275, 1.413)	
Race/Ethnicity	Caucasian	Reference	
	Hispanic	1.106 (1.014, 1.207)	0.023
	Black/African American	1.140 (1.091, 1.191)	< 0.001
	Asian/Pacific Islander	1.004 (0.893, 1.129)	0.947
	Native American	1.145 (0.995, 1.317)	0.059
	Other/Unknown	1.112 (1.061, 1.166)	< 0.001

Previous ED visits (last 6mo)	-	0.621 (0.595, 0.649)	< 0.001
Maximum previous length of stay (last 6mo)	-	1.042 (1.040, 1.044)	< 0.001
Free-standing pediatric hospital	Yes	1.431 (0.659, 3.107)	0.364
EmergentAdmission	Yes	1.672 (1.598, 1.749)	< 0.001
ReadmissionHistory	Yes	1.179 (1.124, 1.236)	< 0.001
Other sepsis (A41)	Yes	1.599 (1.449, 1.764)	< 0.001
Other bacterial agents as the cause of diseases classified elsewhere (B96)	Yes	1.912 (1.740, 2.100)	< 0.001
Lymphoid leukemia (C91)	Yes	0.562 (0.494, 0.640)	< 0.001
Type 1 diabetes mellitus (E10)	Yes	0.279 (0.211, 0.369)	< 0.001
Cerebral palsy (G80)	Yes	1.314 (1.156, 1.494)	< 0.001
Other disorders of brain (G93)	Yes	1.224 (1.119, 1.340)	< 0.001
Suppurative and unspecified otitis media (H66)	Yes	0.522 (0.422, 0.645)	< 0.001
Bacterial pneumonia, not elsewhere classified (J15)	Yes	1.867 (1.664, 2.094)	< 0.001

Pneumonia, unspecified organism (J18)	Yes	0.796 (0.727, 0.872)	< 0.001
Disorders of newborn related to short gestation and low birth weight, not elsewhere classified (P07)	Yes	5.466 (5.084, 5.877)	< 0.001
Respiratory distress of newborn (P22)	Yes	1.839 (1.687, 2.004)	< 0.001
Other respiratory conditions originating in the perinatal period (P28)	Yes	2.579 (2.368, 2.810)	< 0.001
Cardiovascular disorders originating in the perinatal period (P29)	Yes	1.184 (1.071, 1.309)	< 0.001
Other conditions originating in the perinatal period (P96)	Yes	2.412 (2.180, 2.668)	< 0.001
Abnormalities of heart beat (R00)	Yes	1.106 (1.024, 1.195)	0.010
Symptoms and signs concerning food and fluid intake (R63)	Yes	1.630 (1.525, 1.742)	< 0.001
Surgical operation and other surgical procedures as the cause of abnormal reaction of the patient, or of later complication, without mention of misadventure at the time of the procedure (Y83)	Yes	1.227 (1.091, 1.380)	< 0.001

Encounter for observation and evaluation of newborn for suspected diseases and conditions ruled out (Z05)	Yes	1.239 (1.126, 1.363)	< 0.001
Encounter for immunization (Z23)	Yes	1.133 (1.029, 1.248)	0.011
Family history of certain disabilities and chronic diseases (leading to disablement) (Z82)	Yes	0.641 (0.551, 0.745)	< 0.001
Personal history of other diseases and conditions (Z87)	Yes	0.816 (0.734, 0.907)	< 0.001
Other postprocedural states (Z98)	Yes	0.698 (0.632, 0.770)	< 0.001
Pervasive developmental disorders (F84)	Yes	0.577 (0.373, 0.893)	0.013
Sleep disorders (G47)	Yes	0.989 (0.842, 1.162)	0.891
Other functional intestinal disorders (K59)	Yes	1.028 (0.895, 1.181)	0.699
Reaction to severe stress, and adjustment disorders (F43)	Yes	1.354 (1.157, 1.585)	< 0.001
Streptococcus, Staphylococcus, and Enterococcus as the cause of diseases classified elsewhere (B95)	Yes	1.031 (1.017, 1.046)	< 0.001
Other anemias (D64)	Yes	0.965 (0.953, 0.978)	< 0.001

Major depressive disorder, single episode (F32)	Yes	0.946 (0.918, 0.976)	< 0.001
Pervasive developmental disorders (F84)	Yes	1.063 (1.026, 1.102)	< 0.001
Attention-deficit hyperactivity disorders (F90)	Yes	0.934 (0.909, 0.960)	< 0.001
Sleep disorders (G47)	Yes	1.031 (1.014, 1.047)	< 0.001
Asthma (J45)	Yes	1.052 (1.036, 1.068)	< 0.001
Other functional intestinal disorders (K59)	Yes	1.032 (1.020, 1.044)	< 0.001
Obstructive and reflux uropathy (N13)	Yes	0.947 (0.921, 0.974)	< 0.001
Other disorders of urinary system (N39)	Yes	1.034 (1.016, 1.052)	< 0.001
Convulsions, not elsewhere classified (R56)	Yes	0.956 (0.936, 0.977)	< 0.001
Outcome of delivery (Z37)	Yes	0.904 (0.873, 0.935)	< 0.001
Viral and other specified intestinal infections (A08)	Present	1.096 (1.064, 1.128)	< 0.001
Viral agents as the cause of diseases classified elsewhere (B97)	Present	1.044 (1.031, 1.058)	< 0.001
Sickle-cell disorders (D57)	Present	1.042 (1.023, 1.062)	< 0.001

Neutropenia (D70)	Present	1.035 (1.023, 1.047)	< 0.001
Volume depletion (E86)	Present	1.048 (1.038, 1.058)	< 0.001
Other disorders of fluid, electrolyte and acid-base balance (E87)	Present	0.980 (0.975, 0.986)	< 0.001
Reaction to severe stress, and adjustment disorders (F43)	Present	1.032 (1.018, 1.047)	< 0.001
Conduct disorders (F91)	Present	1.066 (1.044, 1.089)	< 0.001
Epilepsy and recurrent seizures (G40)	Present	1.026 (1.017, 1.035)	< 0.001
Essential (primary) hypertension (I10)	Present	0.969 (0.959, 0.978)	< 0.001
Acute upper respiratory infections of multiple and unspecified sites (J06)	Present	1.047 (1.029, 1.065)	< 0.001
Acute bronchiolitis (J21)	Present	1.086 (1.071, 1.101)	< 0.001
Asthma (J45)	Present	1.026 (1.017, 1.034)	< 0.001
Respiratory failure, not elsewhere classified (J96)	Present	0.977 (0.970, 0.983)	< 0.001
Other respiratory disorders (J98)	Present	0.986 (0.979, 0.994)	< 0.001

Gastro-esophageal reflux disease (K21)	Present	1.025 (1.016, 1.033)	< 0.001
Acute appendicitis (K35)	Present	1.076 (1.052, 1.100)	< 0.001
Other functional intestinal disorders (K59)	Present	0.987 (0.980, 0.994)	< 0.001
Cellulitis and acute lymphangitis (L03)	Present	1.065 (1.050, 1.080)	< 0.001
Scoliosis (M41)	Present	0.964 (0.951, 0.976)	< 0.001
Disorders of newborn related to long gestation and high birth weight (P08)	Present	1.124 (1.080, 1.169)	< 0.001
Neonatal jaundice from other and unspecified causes (P59)	Present	1.049 (1.035, 1.063)	< 0.001
Transitory disorders of carbohydrate metabolism specific to newborn (P70)	Present	1.044 (1.022, 1.067)	< 0.001
Feeding problems of newborn (P92)	Present	1.076 (1.059, 1.093)	< 0.001
Congenital malformations of cardiac septa (Q21)	Present	0.955 (0.949, 0.962)	< 0.001
Congenital malformations of great arteries (Q25)	Present	0.965 (0.957, 0.974)	< 0.001
Abnormalities of breathing (R06)	Present	1.043 (1.036, 1.051)	< 0.001

Other symptoms and signs involving the circulatory and respiratory system (R09)	Present	1.022 (1.013, 1.032)	< 0.001
Abdominal and pelvic pain (R10)	Present	1.035 (1.023, 1.047)	< 0.001
Nausea and vomiting (R11)	Present	1.028 (1.019, 1.036)	< 0.001
Symptoms and signs involving emotional state (R45)	Present	1.052 (1.035, 1.069)	< 0.001
Fever of other and unknown origin (R50)	Present	1.036 (1.028, 1.043)	< 0.001
Convulsions, not elsewhere classified (R56)	Present	1.053 (1.041, 1.065)	< 0.001
Lack of expected normal physiological development in childhood and adults (R62)	Present	1.023 (1.015, 1.032)	< 0.001
Place of occurrence of the external cause (Y92)	Present	1.034 (1.023, 1.044)	< 0.001
Outcome of delivery (Z37)	Present	1.047 (1.019, 1.076)	< 0.001
Liveborn infants according to place of birth and type of delivery (Z38)	Present	1.369 (1.342, 1.395)	< 0.001
Encounter for other aftercare and medical care (Z51)	Present	1.048 (1.039, 1.057)	< 0.001
Personal risk factors, not elsewhere classified (Z91)	Present	1.024 (1.012, 1.036)	< 0.001

Artificial opening status (Z93)	Present	1.024 (1.014, 1.033)	< 0.001

Table 4.8: Statistical interactions with age

Variable	**Regression coefficient**	**p value**
Main effects of the interactions		
Age	-0.042	
Streptococcus, Staphylococcus, and Enterococcus as the cause of diseases classified elsewhere (B95)	0.342	
Other anemias (D64)	0.371	
Major depressive disorder, single episode (F32)	1.418	
Pervasive developmental disorders (F84)	-0.549	< 0.001
Attention-deficit hyperactivity disorders (F90)	1.035	
Sleep disorders (G47)	-0.011	
Asthma (J45)	-1.437	
Other functional intestinal disorders (K59)	0.027	
Obstructive and reflux uropathy (N13)	-0.282	
Other disorders of urinary system (N39)	-0.291	
Convulsions, not elsewhere classified (R56)	-0.714	
Outcome of delivery (Z37)	-0.528	
Interaction terms with age		
Streptococcus, Staphylococcus, and Enterococcus as the cause of diseases classified elsewhere (B95)	0.031	
Other anemias (D64)	-0.035	

< 0.001

Major depressive disorder, single episode (F32)	-0.055
Pervasive developmental disorders (F84)	0.061
Attention-deficit hyperactivity disorders (F90)	-0.068
Sleep disorders (G47)	0.030
Asthma (J45)	0.051
Other functional intestinal disorders (K59)	0.031
Obstructive and reflux uropathy (N13)	-0.054
Other disorders of urinary system (N39)	0.033
Convulsions, not elsewhere classified (R56)	-0.045
Outcome of delivery (Z37)	-0.101

Table 4.9: Statistical interactions with the number of medications

Variable	**Regression coefficient**	**p value**
Main effects of the interactions		
Number of medications	0.117	< 0.001
Viral and other specified intestinal infections (A08)	-1.117	
Viral agents as the cause of diseases classified elsewhere (B97)	-0.549	
Sickle-cell disorders (D57)	-1.054	
Neutropenia (D70)	0.199	
Volume depletion (E86)	-0.964	
Other disorders of fluid, electrolyte and acid-base balance (E87)	0.592	
Reaction to severe stress, and adjustment disorders (F43)	0.303	
Conduct disorders (F91)	0.888	
Epilepsy and recurrent seizures (G40)	-0.555	
Essential (primary) hypertension (I10)	0.417	
Acute upper respiratory infections of multiple and unspecified sites (J06)	-0.797	
Acute bronchiolitis (J21)	-1.273	
Asthma (J45)	-1.437	
Respiratory failure, not elsewhere classified (J96)	0.735	
Other respiratory disorders (J98)	0.348	

Gastro-esophageal reflux disease (K21)	-0.157	0.005
Acute appendicitis (K35)	-1.611	< 0.001
Other functional intestinal disorders (K59)	0.027	0.699
Cellulitis and acute lymphangitis (L03)	-1.295	< 0.001
Scoliosis (M41)	-0.034	0.787
Disorders of newborn related to long gestation and high birth weight (P08)	-1.213	< 0.001
Neonatal jaundice from other and unspecified causes (P59)	-0.039	0.421
Transitory disorders of carbohydrate metabolism specific to newborn (P70)	-0.187	0.017
Feeding problems of newborn (P92)	0.422	< 0.001
Congenital malformations of cardiac septa (Q21)	0.405	
Congenital malformations of great arteries (Q25)	0.57	
Abnormalities of breathing (R06)	-0.676	
Other symptoms and signs involving the circulatory and respiratory system (R09)	-0.322	
Abdominal and pelvic pain (R10)	-0.903	
Nausea and vomiting (R11)	-0.519	
Symptoms and signs involving emotional state (R45)	-0.387	
Fever of other and unknown origin (R50)	-0.773	
Convulsions, not elsewhere classified (R56)	-0.714	

Lack of expected normal physiological development in childhood and adults (R62)	-0.124	0.037
Place of occurrence of the external cause (Y92)	-0.764	< 0.001
Outcome of delivery (Z37)	-0.528	
Liveborn infants according to place of birth and type of delivery (Z38)	-1.577	
Encounter for other aftercare and medical care (Z51)	-0.31	
Personal risk factors, not elsewhere classified (Z91)	-0.242	0.002
Artificial opening status (Z93)	-0.241	< 0.001
Interaction terms with Number of medications		
Viral and other specified intestinal infections (A08)	0.091	
Viral agents as the cause of diseases classified elsewhere (B97)	0.043	
Sickle-cell disorders (D57)	0.041	
Neutropenia (D70)	0.034	
Volume depletion (E86)	0.047	
Other disorders of fluid, electrolyte and acid-base balance (E87)	-0.02	
Reaction to severe stress, and adjustment disorders (F43)	0.032	
Conduct disorders (F91)	0.064	
Epilepsy and recurrent seizures (G40)	0.026	
Essential (primary) hypertension (I10)	-0.032	

Acute upper respiratory infections of multiple and unspecified sites (J06)	0.046
Acute bronchiolitis (J21)	0.082
Asthma (J45)	0.025
Respiratory failure, not elsewhere classified (J96)	-0.024
Other respiratory disorders (J98)	-0.014
Gastro-esophageal reflux disease (K21)	0.024
Acute appendicitis (K35)	0.073
Other functional intestinal disorders (K59)	-0.013
Cellulitis and acute lymphangitis (L03)	0.063
Scoliosis (M41)	-0.037
Disorders of newborn related to long gestation and high birth weight (P08)	0.117
Neonatal jaundice from other and unspecified causes (P59)	0.048
Transitory disorders of carbohydrate metabolism specific to newborn (P70)	0.043
Feeding problems of newborn (P92)	0.073
Congenital malformations of cardiac septa (Q21)	-0.046
Congenital malformations of great arteries (Q25)	-0.036
Abnormalities of breathing (R06)	0.042
Other symptoms and signs involving the circulatory and respiratory system (R09)	0.022
Abdominal and pelvic pain (R10)	0.034
Nausea and vomiting (R11)	0.027

Symptoms and signs involving emotional state (R45)	0.051
Fever of other and unknown origin (R50)	0.035
Convulsions, not elsewhere classified (R56)	0.052
Lack of expected normal physiological development in childhood and adults (R62)	0.023
Place of occurrence of the external cause (Y92)	0.033
Outcome of delivery (Z37)	0.046
Liveborn infants according to place of birth and type of delivery (Z38)	0.314
Encounter for other aftercare and medical care (Z51)	0.047
Personal risk factors, not elsewhere classified (Z91)	0.023
Artificial opening status (Z93)	0.023

References

[1] J. D. Cryer and N. Kellet, *Time series analysis* (Springer, 1991).

[2] S. T. Rachev, S. Mittnik, F. J. Fabozzi, S. M. Focardi, and T. Jašić, *Financial econometrics: from basics to advanced modeling techniques*, Vol. 150 (John Wiley & Sons, 2007).

[3] J. Lindberg, Autoregressive conditional density (2016).

[4] R. Kan and X. Wang, On the distribution of the sample autocorrelation coefficients, Journal of Econometrics **154**, 101–121 (2010).

[5] T. J. Fisher, Testing adequacy of arma models using a weighted portmanteau test on the residual autocorrelations, Contributed Paper **327**, 2011 (2011).

[6] G. E. Box and D. A. Pierce, Distribution of residual autocorrelations in autoregressive-integrated moving average time series models, Journal of the American statistical Association **65**, 1509–1526 (1970).

[7] G. M. Ljung and G. E. Box, On a measure of lack of fit in time series models, Biometrika **65**, 297–303 (1978).

[8] M. A. Arranz, Portmanteau test statistics in time series, Time Orientated Language , 1–8 (2005).

[9] Y. Chen, Ting,(2002)" on the robustness of ljung- box and mcleod- li q tests: A simulation study.", Economics Bulletin **3**, 1–10 (2002).

[10] A. C. Monti, A proposal for a residual autocorrelation test in linear models, Biometrika **81**, 776–780 (1994).

[11] D. Peña and J. Rodríguez, A powerful portmanteau test of lack of fit for time series, Journal of the American Statistical Association **97**, 601–610 (2002).

[12] J.-W. Lin and A. I. McLeod, Improved peña–rodriguez portmanteau test, Computational Statistics & Data Analysis **51**, 1731–1738 (2006).

[13] S. K. Safi and A. A. Al-Reqep, Comparative study of portmanteau tests for the residuals autocorrelation in arma models, Comparative study of portmanteau tests for the residuals autocorrelation in ARMA models **2** (2014).

[14] G. Barnard, A new test for 2× 2 tables, Nature **156**, 177 (1945).

[15] R. A. Fisher, On the interpretation of χ 2 from contingency tables, and the calculation of p, Journal of the Royal Statistical Society **85**, 87–94 (1922).

[16] R. Boschloo, Raised conditional level of significance for the 2× 2-table when testing the equality of two probabilities, Statistica Neerlandica **24**, 1–9 (1970).

[17] L. Ehwerhemuepha, H. Sok, and C. Rakovski, A more powerful unconditional exact test of homogeneity for 2× c contingency table analysis, Journal of Applied Statistics , 1–11 (2019).

[18] A. L. Byrne, B. J. Marais, C. D. Mitnick, L. Lecca, and G. B. Marks, Tuberculosis and chronic respiratory disease: a systematic review, International Journal of Infectious Diseases **32**, 138–146 (2015).

[19] S. D. Kumbhare, T. Beiko, S. R. Wilcox, and C. Strange, Characteristics of copd patients using united states emergency care or hospitalization, Chronic Obstructive Pulmonary Diseases **3**, 539 (2016).

[20] K. F. Rabe, S. Hurd, A. Anzueto, P. J. Barnes, S. A. Buist, P. Calverley, Y. Fukuchi, C. Jenkins, R. Rodriguez-Roisin, C. Van Weel, *et al.*, Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: Gold executive summary, American journal of respiratory and critical care medicine **176**, 532–555 (2007).

[21] H. Wang, M. Naghavi, C. Allen, R. M. Barber, Z. A. Bhutta, A. Carter, D. C. Casey, F. J. Charlson, A. Z. Chen, M. M. Coates, and Others, Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015, The lancet **388**, 1459–1544 (2016).

[22] A. A. Cruz, *Global surveillance, prevention and control of chronic respiratory diseases: a comprehensive approach* (World Health Organization, 2007).

[23] P. G. J. Burney, J. Patel, R. Newson, C. Minelli, and M. Naghavi, Global and regional trends in COPD mortality, 1990–2010, European Respiratory Journal **45**, 1239–1247 (2015).

[24] M. I. Asher and P. Ellwood, The global asthma report 2014, (2014).

[25] N. Pearce, N. A\"\it-Khaled, R. Beasley, J. Mallol, U. Keil, E. Mitchell, and C. Robertson, Worldwide trends in the prevalence of asthma symptoms: phase III of the International Study of Asthma and Allergies in Childhood (ISAAC), Thorax **62**, 758–766 (2007).

[26] W. H. Organization, *Global Asthma report 2013* (World Health Organization, 2013).

[27] G. D'Amato, C. Vitale, A. Molino, A. Stanziola, A. Sanduzzi, A. Vatrella, M. Mormile, M. Lanza, G. Calabrese, L. Antonicelli, *et al.*, Asthma-related deaths, Multidisciplinary respiratory medicine **11**, 37 (2016).

[28] N. I. of Health *et al.*, Global initiative for asthma, Global strategy for asthma management and prevention. NHLBI/WHO work shop report (1995).

[29] G. A. Network *et al.*, The global asthma report 2014, Auckland, New Zealand **769**, 28–36 (2014).

[30] W. H. Organization, *Global Asthma report 2014* (World Health Organization, 2014).

[31] T. Nurmagambetov, R. Kuwahara, and P. Garbe, The economic burden of asthma in the United States, 2008–2013, Annals of the American Thoracic Society **15**, 348–356 (2018).

[32] D. Beran, H. J. Zar, C. Perrin, A. M. Menezes, P. Burney, and Others, Burden of asthma and chronic obstructive pulmonary disease and access to essential medicines in low-income and middle-income countries, The Lancet Respiratory Medicine **3**, 159–170 (2015).

[33] C. Troeger, B. Blacker, I. A. Khalil, P. C. Rao, J. Cao, S. R. M. Zimsen, S. B. Albertson, A. Deshpande, T. Farag, Z. Abebe, and Others, Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016, The Lancet infectious diseases **18**, 1191–1210 (2018).

[34] E. Keeler, M. D. Perkins, P. Small, C. Hanson, S. Reed, J. Cunningham, J. E. Aledort, L. Hillborne, M. E. Rafael, F. Girosi, and Others, Reducing the global burden of tuberculosis: the contribution of improved diagnostics, Nature **444**, 49–57 (2006).

[35] J. Ferlay, I. Soerjomataram, M. Ervik, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. M. Parkin, D. Forman, and F. Bray, Cancer incidence and mortality worldwide: IARC CancerBase No. 11 Lyon, France: International Agency for Research on Cancer (2013).

[36] S. M. Jeon, J.-W. Kwon, S. H. Choi, and H.-Y. Park, Economic burden of lung cancer: A retrospective cohort study in South Korea, 2002-2015, PloS one **14** (2019).

[37] J. D. Ribeiro and G. B. Fischer, Chronic obstructive pulmonary diseases in children, Jornal de pediatria **91**, S11—-S25 (2015).

[38] C. Svanes, J. Sunyer, E. Plana, S. Dharmage, J. Heinrich, D. Jarvis, R. de Marco, D. Norbäck, C. Raherison, S. Villani, and Others, Early life origins of chronic obstructive pulmonary disease, Thorax **65**, 14–20 (2010).

[39] D. J. Barker and C. Osmond, Childhood respiratory infection and adult chronic bronchitis in England and Wales., Br Med J (Clin Res Ed) **293**, 1271–1275 (1986).

[40] N. Al-Akour and Y. S. Khader, Having a child with asthma—Quality of life for Jordanian parents, International journal of nursing practice **15**, 574–579 (2009).

[41] I. Rydström, A.-C. Dalheim-Englund, B. Holritz-Rasmussen, C. Möller, and P.-O. Sandman, Asthma–quality of life for Swedish children, Journal of clinical nursing **14**, 739–749 (2005).

[42] L. J. Akinbami, J. E. Moorman, and X. Liu, Asthma prevalence, health care use, and mortality; United States, 2005-2009, (2011).

[43] A. J. Janse, G. Sinnema, C. Uiterwaal, J. L. L. Kimpen, and R. Gemke, Quality of life in chronic illness: perceptions of parents and paediatricians, Archives of Disease in Childhood **90**, 486–491 (2005).

[44] F. Abuekteish, R. Alwash, M. Hassan, and A. S. Daoud, Prevalence of asthma and wheeze in primary school children in northern Jordan, Annals of tropical paediatrics **16**, 227–231 (1996).

[45] M. L. Burr, B. K. Butland, S. King, and E. Vaughan-Williams, Changes in asthma prevalence: two surveys 15 years apart., Archives of disease in childhood **64**, 1452–1456 (1989).

[46] P. G. Burney, S. Chinn, and R. J. Rona, Has the prevalence of asthma increased in children? Evidence from the national study of health and growth 1973-86., Bmj **300**, 1306–1310 (1990).

[47] H. R. Anderson, P. A. Bailey, J. S. Cooper, J. C. Palmer, and S. West, Morbidity and school absence caused by asthma and wheezing illness., Archives of disease in childhood **58**, 777–784 (1983).

[48] A. H. Liu, A. W. Gilsenan, R. H. Stanford, W. Lincourt, R. Ziemiecki, and H. Ortega, Status of asthma control in pediatric primary care: results from the pediatric Asthma Control Characteristics and Prevalence Survey Study (ACCESS), The Journal of pediatrics **157**, 276–281 (2010).

[49] H. J. Zar and T. W. Ferkol, The global burden of respiratory disease—impact on child health, Pediatric pulmonology **49**, 430–434 (2014).

[50] C. K. Lai, R. Beasley, J. Crane, S. Foliaki, J. Shah, and S. Weiland, International Study of Asthma and Allergies in Childhood Phase Three Study Group. Global variation in the prevalence and severity of asthma symptoms: phase three of the International Study of Asthma and Allergies in Childhood (ISAAC), Thorax **64**, 476–483 (2009).

[51] P. Eamranond and E. Jaramillo, Tuberculosis in children: reassessing the need for improved diagnosis in global control strategies, The International Journal of Tuberculosis and Lung Disease **5**, 594–603 (2001).

[52] A. Kochi, The global tuberculosis situation and the new control strategy of the World Health Organization (2001).

[53] M. C. Raviglione, D. E. Snider, and A. Kochi, Global epidemiology of tuberculosis: morbidity and mortality of a worldwide epidemic, Jama **273**, 220–226 (1995).

[54] A. Van Rie, N. Beyers, R. P. Gie, M. Kunneke, L. Zietsman, and P. R. Donald, Childhood tuberculosis in an urban population in South Africa: burden and risk factor, Archives of disease in childhood **80**, 433–437 (1999).

[55] C. M. Perez-Velez and B. J. Marais, Tuberculosis in children, New England Journal of Medicine **367**, 348–361 (2012).

[56] L. J. Nelson and C. D. Wells, Tuberculosis in children: considerations for children from developing countries, in *Seminars in pediatric infectious diseases*, Vol. 15 (Elsevier, 2004) pp. 150–154.

[57] B. Kampmann, Paediatric Tuberculosis, Lancet Infect Dis **8**, 498–510 (2008).

[58] W. H. Organization and Others, *Childhood tuberculosis and BCG vaccine* (World Health Organization, 1989).

[59] E. L. Corbett, C. J. Watt, N. Walker, D. Maher, B. G. Williams, M. C. Raviglione, and C. Dye, The growing burden of tuberculosis: global trends and interactions with the HIV epidemic, Archives of internal medicine **163**, 1009–1021 (2003).

[60] H. E. Jenkins, Global burden of childhood tuberculosis, Pneumonia **8**, 24 (2016).

[61] N. C. V. De Sales, C. Moss, I. Brooks, K. Hood, and K. Nelson, Unscheduled returns to the emergency department, Australian Nursing and Midwifery Journal **22**, 41 (2014).

[62] G. L. Goh, P. Huang, M. C. P. Kong, S.-P. Chew, and S. Ganapathy, Unplanned reattendances at the paediatric emergency department within 72 hours: a one-year experience in KKH, Singapore medical journal **57**, 307 (2016).

[63] C. S. Cho, D. J. Shapiro, M. D. Cabana, J. H. Maselli, and A. L. Hersh, A national depiction of children with return visits to the emergency department within 72 hours, 2001–2007, Pediatric emergency care **28**, 606–610 (2012).

[64] F. Meng, K. L. Teow, K. W. S. Teo, C. K. Ooi, and S. Y. Tay, Predicting 72-hour reattendance in emergency departments using discriminant analysis via mixed integer programming with electronic medical records, Journal of Industrial & Management Optimization **15**, 947–962 (2019).

[65] I. Imsuwan, Characteristics of unscheduled emergency department return visit patients within 48 hours in Thammasat University Hospital, J Med Assoc Thai **94**, S73—-80 (2011).

[66] A. E. Ahmed, B. I. ALMuqbil, M. N. Alrajhi, H. R. Almazroa, D. A. AlBuraikan, M. A. Albaijan, M. Nasim, M. A. Alsalamah, D. K. McClish, and A.-J. Hamdan, Emergency department 72-hour revisits among children with chronic diseases: a Saudi Arabian study, BMC pediatrics **18**, 205 (2018).

[67] M. C. Monuteaux, L. Lee, and E. Fleegler, Children injured by violence in the United States: emergency department utilization, 2000–2008, Academic emergency medicine **19**, 535–540 (2012).

[68] S.-F. Sung, K. E. Liu, S. C.-C. Chen, C.-L. Lo, K.-C. Lin, and Y.-H. Hu, Predicting factors and risk stratification for return visits to the emergency department within 72 hours in pediatric patients, Pediatric emergency care **31**, 819–824 (2015).

[69] B. D. Shy, E. Y. Kim, N. G. Genes, T. Lowry, G. T. Loo, U. Hwang, L. D. Richardson, and J. S. Shapiro, Increased identification of emergency department 72-hour returns using multihospital health information exchange, Academic Emergency Medicine **23**, 645–649 (2016).

[70] A. T. Akenroye, C. W. Thurm, M. I. Neuman, E. R. Alpern, G. Srivastava, S. P. Spencer, H. K. Simon, J. Tejedor-Sojo, C. H. Gosdin, E. Brennan, and Others, Prevalence and predictors of return visits to pediatric emergency departments, Journal of hospital medicine **9**, 779–787 (2014).

[71] J. Li, A. De, K. Ketchum, L. Fagnan, D. G. Haxby, and A. Thomas, Antimicrobial prescribing for upper respiratory infections and its effect on return visits, Fam Med **41**, 182–7 (2009).

[72] B. Strack, J. P. DeShazo, C. Gennings, J. L. Olmo, S. Ventura, K. J. Cios, and J. N. Clore, Impact of HbA1c measurement on hospital readmission rates: analysis of 70,000 clinical database patient records, BioMed research international **2014** (2014).

[73] E. R. Mansfield and B. P. Helms, Detecting multicollinearity, The American Statistician **36**, 158–160 (1982).

[74] J. M. Cortina, Interaction, nonlinearity, and multicollinearity: Implications for multiple regression, Journal of Management **19**, 915–922 (1993).

[75] R. M. O'brien, A caution regarding rules of thumb for variance inflation factors, Quality & quantity **41**, 673–690 (2007).

[76] I. Osman, D. Godden, J. Friend, J. Legge, and J. Douglas, Quality of life and hospital readmission in patients with chronic obstructive pulmonary disease., Thorax **52**, 67–71 (1997).

[77] J. Garcia-Aymerich, E. Monso, R. M. Marrades, J. ESCARRABILL, M. A. Felez, J. Sunyer, J. M. Anto, and E. investigators, Risk factors for hospitalization for a chronic obstructive pulmonary disease exacerbation: Efram study, American journal of respiratory and critical care medicine **164**, 1002–1007 (2001).

[78] H. Baek, M. Cho, S. Kim, H. Hwang, M. Song, and S. Yoo, Analysis of length of hospital stay using electronic health records: A statistical and data mining approach, PloS one **13**, e0195901 (2018).

[79] B. A. Marfil-Garza, P. F. Belaunzarán-Zamudio, A. Gulias-Herrero, A. C. Zuñiga, Y. Caro-Vega, D. Kershenobich-Stalnikowitz, and J. Sifuentes-Osornio, Risk factors associated with prolonged hospital length-of-stay: 18-year retrospective study of hospitalizations in a tertiary healthcare center in mexico, PloS one **13**, e0207203 (2018).

[80] K. J. Brasel, H. J. Lim, R. Nirula, and J. A. Weigelt, Length of stay: an appropriate quality measure?, Archives of Surgery **142**, 461–466 (2007).

[81] B. Davis, S. Sullivan, A. Levine, and J. Dallara, Factors affecting ed length-of-stay in surgical critical care patients, The American journal of emergency medicine **13**, 495–500 (1995).

[82] R. Esmaeili, S.-M. Aghili, M. Sedaghat, and M. Afzalimoghaddam, Causes of prolonged emergency department stay; a cross-sectional action research, Advanced journal of emergency medicine **2** (2018).

[83] H. Bueno, J. S. Ross, Y. Wang, J. Chen, M. T. Vidán, S.-L. T. Normand, J. P. Curtis, E. E. Drye, J. H. Lichtman, P. S. Keenan, *et al.*, Trends in length of stay and short-term outcomes among medicare patients hospitalized for heart failure, 1993-2006, Jama **303**, 2141–2147 (2010).

[84] T. Rotter, L. Kinsman, E. L. James, A. Machotta, H. Gothe, J. Willis, P. Snow, and J. Kugler, Clinical pathways: effects on professional practice, patient outcomes, length of stay and hospital costs, Cochrane database of systematic reviews (2010).

[85] M. B. Rosenman, T. Vik, S. L. Hui, and P. P. Breitfeld, Hospital resource utilization in childhood cancer, Journal of pediatric hematology/oncology **27**, 295–300 (2005).

[86] C. Morley, M. Unwin, G. M. Peterson, J. Stankovich, and L. Kinsman, Emergency department crowding: A systematic review of causes, consequences and solutions, PloS one **13**, e0203316 (2018).

[87] K. G. Negari, K. M. Rodamo, A. T. Hirigo, *et al.*, Factors associated with the length of stay in emergency departments in southern-ethiopia, BMC research notes **12**, 239 (2019).

[88] A. Herring, A. Wilper, D. U. Himmelstein, S. Woolhandler, J. A. Espinola, D. F. Brown, and C. A. Camargo Jr, Increasing length of stay among adult visits to us emergency departments, 2001–2005, Academic Emergency Medicine **16**, 609–616 (2009).

[89] P. Yoon, I. Steiner, and G. Reinhardt, Analysis of factors influencing length of stay in the emergency department, Canadian Journal of Emergency Medicine **5**, 155–161 (2003).

[90] D. Gupta, P. G. Vashi, C. A. Lammersfeld, and D. P. Braun, Role of nutritional status in predicting the length of stay in cancer: a systematic review of the epidemiological literature, Annals of Nutrition and Metabolism **59**, 96–106 (2011).

[91] R. G. Rosa and L. Z. Goldani, Factors associated with hospital length of stay among cancer patients with febrile neutropenia, PLoS One **9**, e108969 (2014).

[92] R. Berardi, M. Caramanti, M. Castagnani, S. Guglielmi, F. Marcucci, A. Savini, F. Morgese, S. Rinaldi, C. Ferrini, M. Tiberi, *et al.*, Hyponatremia is a predictor of hospital length and cost of stay and outcome in cancer patients, Supportive Care in Cancer **23**, 3095–3101 (2015).

[93] T. Rotter, J. Kugler, R. Koch, H. Gothe, S. Twork, J. M. van Oostrum, and E. W. Steyerberg, A systematic review and meta-analysis of the effects of clinical pathways on length of stay, hospital costs and patient outcomes, BMC health services research **8**, 265 (2008).

[94] E. K. Goodman, A. F. Reilly, B. T. Fisher, J. Fitzgerald, Y. Li, A. E. Seif, Y.-S. Huang, R. Bagatell, and R. Aplenc, Association of weekend admission with hospital length of stay, time to chemotherapy, and risk for respiratory failure in pediatric patients with newly diagnosed leukemia at freestanding us children's hospitals, JAMA pediatrics **168**, 925–931 (2014).

[95] F. Feliciana Silva, G. Macedo da Silva Bonfante, I. A. Reis, H. André da Rocha, A. Pereira Lana, and M. Leal Cherchiglia, Hospitalizations and length of stay of cancer patients: A cohort study in the brazilian public health system, PloS one **15**, e0233293 (2020).

[96] L. Ehwerhemuepha, G. Gasperino, N. Bischoff, S. Taraman, A. Chang, and W. Feaster, Healthedatalab–a cloud computing solution for data science and advanced analytics in healthcare with application to predicting multi-center pediatric readmissions, BMC medical informatics and decision making **20**, 1–12 (2020).

[97] D. A. Zighed, S. Lallich, and F. Muhlenbach, A statistical approach to class separability, Applied Stochastic Models in Business and Industry **21**, 187–197 (2005).

[98] R. J. Boik, Interactions, partial interactions, and interaction contrasts in the analysis of variance., Psychological Bulletin **86**, 1084 (1979).

[99] D. Hedeker, A mixed-effects multinomial logistic regression model, Statistics in medicine **22**, 1433–1446 (2003).

[100] R. C. Team *et al.*, R: A language and environment for statistical computing (2013).

[101] J. F. Box, Ra fisher and the design of experiments, 1922–1926, The American Statistician **34**, 1–7 (1980).